Warped Hypertime Representations for Long-term Autonomy of Mobile Robots

Tomáš Krajník, Tomáš Vintr, Sergi Molina, Jaime P. Fentanes, Gregorz Cielniak, Tom Duckett

Abstract—This paper presents a novel method for introducing time into discrete and continuous spatial representations used in mobile robotics, by modelling long-term, pseudo-periodic variations caused by human activities. Unlike previous approaches, the proposed method does not treat time and space separately, and its continuous nature respects both the temporal and spatial continuity of the modeled phenomena. The method extends the given spatial model with a set of wrapped dimensions that represent the periodicities of observed changes. By performing clustering over this extended representation, we obtain a model that allows us to predict future states of both discrete and continuous spatial representations. We apply the proposed algorithm to several long-term datasets and show that the method enables a robot to predict future states of representations with different dimensions. The experiments further show that the method achieves more accurate predictions than the previous state of the art.

I. INTRODUCTION

Advances in autonomous robotics are gradually enabling deployment of robots in human-populated environments [1]. Human activity tends to cause changes to the environments it takes place in, and the mobile robots that share these environments need to be able to cope with such never-ending changes. However, merely coping is not enough – ideally, a mobile robot should not only be able to comprehend the environment structure, but also understand how it changes over time. Many authors have shown that even simple models of the environment that adapt to changes improve the overall ability of mobile robots to operate over longer time periods [2], [3], [4], [5], [6]. Moreover, the ability of long-term autonomous operation improves the chances of observing the temporal changes, and thus to extract more valuable information about the environment dynamics, resulting in more accurate spatio-temporal models [1].

Our approach proposes to extend spatial representations by adding several dimensions representing time, with each pair of temporal dimensions representing a given periodicity observed in the gathered data. In particular, we employ the Frequency Map Enhancement [5] concept to identify (temporally) periodic patterns in the data gathered. This approach represents periodic processes in the environment using Fourier analysis, and the resulting spectral models obtained from long-term experience can be used to predict future environment states. Then, we transform each time periodicity into a pair of dimensions that form a circle in 2d space and add these dimensions to the vectors that represent the spatial aspects of the modelled phenomena. The model itself is then built using traditional techniques like clustering or expectation-maximisation over the warped space-hypertime representation. The resulting multi-modal model represents both the structure of the space and temporal patterns of the changes or events. We hypothesize that since the proposed model respects the spatio-temporal continuity of the modelled phenomena, it provides more accurate predictions than models that partition the modeled space into discrete elements. In this paper, we provide a description of the proposed method, and provide experimental evidence of its capability to efficiently represent spatio-temporal data and to predict future states of the environment. Unlike the previous works [3], [8], [4], [5], which can only introduce time into models that represent the environment by a discrete set of binary states, such as visibility of landmarks or cell occupancy in grids, our method is able to work with continuous and higher-dimensional variables, e.g. robot velocities, object positions, pedestrian flows, etc. Moreover, the method explicitly represents and predicts not only the value of a given state, but also its probabilistic distribution.
at a particular time and location, which can be useful for task scheduling and planning [9].

Our experiments, based on real world data gathered over several weeks, confirm that the method achieved more accurate predictions than both static models and models that aim to represent time over a discretised space only.

II. RELATED WORK

In mobile robotics, the effects of environment variations were studied mainly from the perspective of localisation and mapping, because neglecting the environment change gradually deteriorates the ability of the robot to determine its position reliably and accurately. Assuming that the world is non-stationary, the authors of [2], [10], [11], [12], [6] proposed approaches that create, refine and update world models in a continuous manner. Furthermore, Ambrus et al. [13] demonstrated that the ability of continuous remapping not only allows to refine models of the static environment structure, but also opens up the possibility to learn object models from the spatial changes observed [14].

Unlike the aforementioned works, which focused on the spatial structure and appearance aspects of the changes observed, other authors [2], [15], [16], [3], [5] focused on modelling the temporal aspects. For example, [2] and [15] represent the environment dynamics by multiple temporal models with different timescales, where the best map for localisation is chosen by its consistency with current readings. Dayoub et al. [16] and Rosen et al. [8] used statistical methods to create feature persistence models and reasoned about the stability of the environmental states over time. Tipaldi et al. [3] proposed to represent the occupancy of cells in a traditional occupancy grid with a Hidden Markov Model. Krajnik et al. [5] represent the probability of environment states in the spectral domain, which captures cyclic (daily, weekly, yearly) patterns of environmental changes as well as their persistence.

The aforementioned approaches demonstrated that considering temporal aspects (and especially their persistence and periodicity) in robotic models improves not only mobile robot localisation [5], [3], [2], but also planning [17], [18] and exploration [19]. However, these temporal representations were tailored to model the probability of a single state over time, and thus were applied only to individual components of the discretised models, e.g. cells in an occupancy grid [3], [19], visibility of landmarks [5], traversability of edges in topological maps [17] or human presence in a particular room [18]. Since the spatial interdependence of these components was neglected, the above models were actually considering only temporal and not spatial-temporal relations of the represented environments. This results not only in memory inefficiency (because of the necessity to model a high number of discrete states separately) but also in the inability of the representation to estimate environment states at locations where no measurements were taken, e.g. if a certain cell in an occupancy grid is occluded, its state is unknown even if the neighbouring cells are occupied, because the cell is part of a wall or ground.

Spatio-temporal relations of discrete environment models were investigated in [20], [21], Kuczner et al. [20] 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. A similar approach was proposed in [21], where the direction of traversal over each cell is obtained using an input-output Hidden Markov Model connected to neighboring cells. However, these models represent only local spatial dependencies and suffer from a major disadvantage of the discretised models – memory inefficiency. Therefore, in their latest work, [22] model a given set of spatio-temporal phenomena (the motion of people and wind flow) in a continuous domain, building their model by means of Expectation Maximisation. Moreover, [23] shows how to use this representation for robot motion planning in crowded environments.

O’Callaghan and Ramos [24] also argue in favour of continuous models, showing the advantages of Gaussian Mixture-based representations in terms of memory efficiency and utility for mobile robot navigation. In their latest work, [25] speed up building and updating of the proposed models by using an elegant combination of kernels and optimization methods. The speed-up achieved allows to recompute the model relatively quickly, which keeps the model updated with the changes in the robot’s operational environment.

Unlike the work of Ramos et al. [25], which is aimed primarily at modelling the spatial structure, and [22], which aims to make short-term predictions of the motion of people, our aim is to create universal, spatial-temporal models capable of long-term predictions of various phenomena. Inspired by the ability of the continuous models [25], [22] to represent spatio-temporal phenomena and the predictive power of spectral representations [5], we propose a novel method which allows to introduce the notion of time into state-of-the-art spatial models used in mobile robotics. Unlike our previous work [5], which treats environmental states as independent despite their spatial proximity and is applicable to binary states only, the proposed method can be applied to continuous, multi-dimensional representations, e.g. object positions, movements of people, gas concentration, etc. Moreover, the method explicitly represents and predicts not only the value of a given state, but also its probabilistic distribution at a particular time and location, which can be useful for task scheduling and planning.

III. METHOD

A. Example scenario

As an example, let us consider a robot providing an information terminal service in an office building or a conference guide that has to provide directions to certain events at locations where people commonly gather before the events. To provide the service efficiently, the robot has to position itself close to areas with a high level of pedestrian traffic. However, to avoid causing nuisance to people and improve robot performance, the robot should arrive at these locations before they become congested. Thus, the robot
has to anticipate the likelihood of occurrence of people at a given time and place based on a model, which is built from data gathered by the robot’s people detection system. Such a system provides the robot with \( x, y \in \mathbb{R}^2 \) positions of the people in the robot’s field of view. To be able to navigate to this spot efficiently, the robot should know how to predict if, and how quickly, it is able to reach the desired destination from the current position. To do so, it has to predict which of the potential paths will be blocked (e.g. due to doors being closed) and how quickly it will be able to traverse them. Furthermore, the robot has to reason about the ability to localise itself reliably along the path it plans to traverse. This means that the robot needs to use past data from its navigation system, including its velocity \( v \in \mathbb{R} \) along a given path and success of the path traversal \( s \in \{0, 1\} \). The number of people \( n \), robot speed \( v \) and path traversal success \( s \) around the position \((x, y)\) might be influenced by time, since people gather in different areas at different times (meeting rooms, cafeterias, etc.), and the robot speed is influenced by the congestion of the corridors it moves through and the doors on the path that are more likely to be open during busy hours than otherwise.

Thus, one can assume that the values of \( n, v \) and \( s \), which form the state of the environment, will appear with a different frequency depending on the time \( t \) and location \( x, y \). In particular, they are likely to be influenced by patterns of human activity, which are strongly influenced by the time of day and the day of the week. Our method provides a unified way to identify the dependence of these variables on time by modelling the spatio-temporal distribution of their occurrences. The problem of finding such distributions is that while the modelled space is constrained, and thus one can gather an arbitrary number of measurements from a given location, time unfolds indeﬁnitely and it is not possible to obtain measurements with the same \( t \), which makes calculation of the temporal density of some phenomena diﬃcult. In our case, we assume that the time domain exhibits certain periodic properties, and project the entire time-line into a multi-dimensional, but constrained warped space, where the notion of event density makes sense.

B. Method Overview

Let us assume that a robot already gathered a training set containing \( l \) measurements of a given phenomenon, obtaining tuples \((a_i, x_i, t_i)\), where \( i \in \{1 \ldots l\} \), the vector \( x_i \) describes the location of the measurement (e.g. position of a detected person or obstacle), \( t_i \) corresponds to the time of the measurement and \( a_i \) represents the measurement’s value, e.g. the number of detected people, likelihood of an obstacle or robot velocity in the vicinity of \((x, t)\).

Our method aims to find a \( p(a|x, t) \), which would represent the conditional probability density function of the variable \( a \) given the position \( x \) and time \( t \). The proposed method is composed of five stages, which are performed in an iterative manner:

1) initialization;
2) spatio-temporal clustering;
3) model error estimation;
4) identification of periodicities;
5) and hypertime space extension.

To initialize the algorithm, we first set an index \( h \), which characterises the number of periodicities taken into the temporal model, to zero and we store all measurements \((a_i, x_i)\) in \( h_x \). In the spatio-temporal clustering stage, we cluster the vectors \((h_x_i)\), obtaining a Gaussian mixture model, which represents the spatio-temporal distribution of the given phenomenon and allows to calculate conditional probability function \( p_h(a|x, t) \). In the model error estimation, we calculate the mean \( \mu \) of \( p(a|x, t) \) for all training samples. Then we calculate the time series \( h\epsilon(t_i) \) as \( h\epsilon(t_i) = \mu - a_i \) and its mean squared value \( E_h \). Then, during the identification of periodicities, we perform spectral analysis of \( h\epsilon(t_i) \), extract the most prominent spectral component, and store its period as \( T_{h+1} \). After that, we perform the hypertime space extension, which extends each vector \( h_x \) by 2 dimensions representing a given periodicity of the temporal domain, i.e.

\[
\begin{align*}
  h+1_x_i &\leftarrow (h_x_i, \cos(2\pi t_i/T_{h+1}), \sin(2\pi t_i/T_{h+1})).
  \end{align*}
\]

Then, we increment \( h \) by one and repeat the steps of spatio-temporal clustering and model error estimation on the now extended vectors \( h+1_x \), obtaining an new error \( E_{h} \). We compare the model error \( E_{h} \) calculated with the error obtained in the previous iteration \( E_{h-1} \) and if \( E_{h} < E_{h-1} \), we proceed with identification of periodicities and hypertime space extension, extending the vector \( h+1_x \) with another two dimensions representing another potential periodicity of the modeled phenomena. In cases where the model error starts to increase, i.e. if \( E_{h} \geq E_{h-1} \), we store the model \( p_{h-1}(a|x, t) \) from the previous iteration as \( p(a|x, t) \) and terminate the method.

The resulting model allows to estimate the likelihood of each value \( a \) of a given phenomena at location \( x \) and time \( t \). In our experiments, we show that the function \( p(a|x, t) \) allows to predict the visibility of image features, door states, robot velocity and number of people occurrences within a given spatio-temporal volume.

C. Spatio-Temporal Clustering

We represent the probability density function \( p(a|x, t) \) by a mixture of Gaussian models in the hypertime space as follows:

\[
  p(a|x, t) = \gamma \sum_{j=1}^{n} w_j u_j(a, x, h^t),
\]

where \( u_j(a, x, h^t) \) is a multivariate Gaussian function of the \( j^{th} \) cluster, \( w_j \) is the cluster weight, \( h^t = (\cos(2\pi t_1/T_{h}), \sin(2\pi t_1/T_{h}), \ldots, \cos(2\pi t_j/T_{h}), \sin(2\pi t_j/T_{h})) \) is the projection of time in the hypertime space and \( \gamma \) is a scaling constant. Calculation of the model, i.e. computation of weights, means and covariances of the Gaussian mixture model is performed by an Expectation Maximisation scheme, discussed in detail in Chapter [IV].
D. Model error estimation

Projecting the linear time $t$ onto the circular hypertime space (or its inverse) inevitably changes the scale of the calculated spatio-temporal density. This is because several time instants $t$ can project into the same area of hypertime. Thus, we first need to determine the scaling factor $\gamma$ in such a way that the mean value of $a$ calculated from the model over the training set vectors $(x_i, t_i)$ is equal to the average value of $a_i$ on the training set:

$$
\gamma = \frac{\sum_{i=1}^{l} a_i}{\sum_{i=1}^{l} \sum_{j=1}^{n} w_j u_j (a, x_i, h t_i) da}
$$

(3)

After calculating the scaling factor $\gamma$, we compute an estimate of $a_i$ at each training test point defined by location $x_i$ and time $t_i$ by calculating the mean $\mu_i$:

$$
\mu_i = \int a p(a, x_i, t_i) da,
$$

(4)

Then, we calculate the error $h \epsilon(t_i)$ as the difference between the mean and the measured values $a_i$:

$$
h \epsilon(t_i) = \mu_i - a_i.
$$

(5)

Finally, we calculate the mean squared error of the current model as

$$
E_h = \sqrt{\sum_{i=1}^{l} h \epsilon_i^2 (t_i)} = \sqrt{\sum_{i=1}^{l} (\mu_i - a_i)^2}.
$$

(6)

We compare the mean squared error $E_h$ with the one calculated in the previous iteration $E_{h-1}$. If $E_h$ is smaller than $E_{h-1}$, then we increase the $h$ by one and perform another iteration of the method.

E. Identification of periodicities and hypertime extension

To identify the periodicities in the error, we use a Fourier-transform scheme. However, since the data collections for the experiments were performed by a system operating in real-world conditions, they were not collected in a (temporally) regular manner. Thus, we process the time series $h \tilde{\epsilon}(t_i)$ by the FreMEn method [5], which is able to find periodicities in non-uniform and sparse data. In particular, we calculate the most prominent periodicity $T_h$ in the error time-series as:

$$
T_h = \arg \max_{t_k} \sum_{i=1}^{l} |(h \tilde{\epsilon}(t_i) - h \hat{\epsilon}) e^{-j 2 \pi t_i / T_h}|.
$$

(7)

where $h \hat{\epsilon}$ is an average error $h \epsilon(t_i)$. After establishing the $T_h$, we extend all vectors of the training set $h^{-1} x(t_i)$ by adding another two components $(\cos(2 \pi t_i / T_h), \sin(2 \pi t_i / T_h))$, i.e.

$$
h x_i \leftarrow (h^{-1} x_i, \cos(2 \pi t_i / T_h), \sin(2 \pi t_i / T_h)).
$$

(8)

Thus, at the start of out method, each vector $0 x_i$ contains only the spatial information, i.e. $0 x_i = (a_i, x_i)$, but at the end, the vector contains $2h$ additional dimensions modeling the periodicities observed in the training data.

IV. Discussion of Clustering

While the hypertime extension, error estimation and periodicity estimation steps of the method are quite straightforward, deterministic and computationally inexpensive, the way we build the model of the probability distribution over the hypertime space is key to the method’s predictive efficiency. The main issue of the hypertime space is its sparsity, because the time, which is linear and one-dimensional, is projected onto a single-dimensional curve lying on a multi-dimensional hypersphere. This causes problems with the numerical stability of many algorithms that we tested. Thus, we dedicated a significant effort into testing various clustering methods, their initialisations and metrics. Since this letter is concerned with the idea of using the hypertime space to allow robots to take into account cyclic environment variations during their long-term operation, we will give a short overview of the methods tested on the scenarios described in Section [2]. A thorough comparison of the aforementioned methods and settings is beyond the scope of this article and will be presented in a separate paper.

We evaluated different clustering methods and metrics to model distributions in the spatio-temporal hyperspace. The Gustafson–Kessel algorithm [26] with fuzzy membership degrees [27], which uses Mahalanobis distance, worked well on a space with two temporal and two spatial dimensions. However, as the number of modelled periodicities (and thus, the number of temporal dimensions) grew, the method became unstable and did not provide meaningful results. This confirms our previous results [28], which showed that Gustafson–Kessel does not achieve good performance in high-dimensional spaces. We also evaluated other distribution modeling methods described in [29]. In particular, we thoroughly evaluated the performance of fuzzy k-means clustering [30] and classical k-means [31]. Surprisingly, the only clustering method able to partition the hypertime-space into meaningful clusters was classical k-means [31]. We also attempted to model the spatio-temporal hyperspace by a mix of Gaussian distributions, determined by an Expectation-Maximisation scheme similar to [25], [22]. While these achieved good results in most cases, they sometimes exhibited numerical instabilities similar to the Gustafson–Kessel algorithm [26]. However, the effects of numerical instability are possible to detect through eigenvalue analysis of the covariance matrices and if needed, the EM can be restarted with different initialization or with the covariances matrices restricted to diagonal-only.

We also evaluated a variety of different metrics, e.g. Euclidean, Minkowski, Chebyshev, etc., as well as their squared and square rooted variants. The Minkowski metric is recommended for clustering in higher dimensions [32], but it performed no better than Euclidean distance when compared on hypertime space. As mentioned earlier, we also evaluated Mahalanobis distance without satisfactory results. Inspired by [33][34], we experimented with the mixtures of cosine distance for hypertimes and the aforementioned metrics for other variables.
Finally, in our experiments we compare two clustering methods, which provided the most satisfactory results, the HyperTime Expectation Maximisation (HyT-EM) and Hyper-Time K-Means (HyT-KM).

The HyT-EM method is based on the Expectation Maximisation scheme implementation from the OpenCV library. As the method requires to specify the number of clusters, we indicate the the method name as HyT-EM\_k, where k is the number of clusters used during the experiments (Section V). As mentioned before, the only problem of the method is its occasional numerical instability, which, if detected, is solved by automated restart with different initial positions of the clusters.

The HyT-KM method is based on the NumPy package, popular for scientific computing in the Python language. HyT-KM first initialises the cluster centres using k-means, while using the cosine distance for temporal and Euclidean distance for spatial dimensions. After initialisation, it calculates the covariace matrices of the clusters and then it proceeds with the standard Expectation Maximisation procedure, while using the cosine distance for temporal dimensions. Unlike HyT-EM, HyT-KM does not require to specify the number of clusters in advance. Instead, the algorithm tries to analyse the temporal structure of the hypertime-space prior to the hypertime expansion step. In particular, it starts with \( n = 1 \) clusters and calculates the sum of amplitudes \( T_{\Sigma}(n) \) and \( T_{\Sigma}(n+1) \) of the frequency spectrum of the error \( h \epsilon(t_i) \)

\[
T_{\Sigma}(n) = \sum_{k=1}^{K} \sum_{i=1}^{d} |(h \epsilon(t_i) - b \epsilon_t)| e^{-j2\pi t_i / T_k}.
\]

If \( T_{\Sigma}(n) > T_{\Sigma}(n+1) \), the model with \( n + 1 \) clusters is stored and the number of clusters \( n \) is increased by 1. If \( T_{\Sigma}(n) \leq T_{\Sigma}(n+1) \), the method simply proceeds with the hypertime expansion step.

V. EXPERIMENTS

The purpose of the experimental evaluation is to assess the predictive capability of the proposed method and its utility for different robotic tasks. The performance of the method is evaluated in four different scenarios, which require predictions of variables of different dimensionalities. The data for these experiments were collected by robotic sensors in real world conditions over periods of several weeks. These scenarios correspond to our original motivational example from Section III-B where we discussed how a long-term operating robot will benefit from the predictive capabilities of models that explicitly represent temporal behaviour of environment states with different dimensions. To evaluate the efficiency of our method, we compare five different temporal models: Mean, which predicts a value as an average of its past measurements, Hist, which divides each day into \( n \) intervals and predicts the given variable as an average in a relevant time of a day, FreMEn\_m, which extracts \( m \) periodic components from the variable’s history and uses these periodicities for prediction, HyT-EM\_k, which uses the expectation-maximisation of \( k \)-component Gaussian Mixture Model over the hypertime space, and finally HyT-KN, as described in Section IV. The experimental evaluation is performed by an automated system [35], which first optimises each method’s parameters (number of intervals \( n \), number of periodicities \( m \), and number of clusters \( k \)) and then runs a series of pairwise t-tests to determine which methods perform statistically significantly better than other ones. To enable the reproducibility of the results, the evaluation system and the datasets are available online [36].

A. Door state

The first scenario concerns a single binary variable, which corresponds to the state of a university office door. The door was continuously observed by an RGB-D camera for 10 weeks to obtain the training set, and for another 10 weeks to obtain 10 testing sets, each one week long. Since the RGB-D data processing was rather simple, the data contains noise, because people moving through the door caused the system to indicate incorrectly that the door was closed.

To compare the efficiency of the predictions, we calculated the mean squared error \( \epsilon \) of the various temporal models’ predictions \( p(t) \) to the ground truth \( s(t) \) as \( \epsilon = \sqrt{T \sum (p(t) - s(t))^2 / |s(t)|} \). The results indicate that both hypertime-based models outperformed the other ones, including FreMEn [5]. Both methods indicated that the most prominent periodicities corresponded to one day, four hours and one week.

B. Topological localisation

In this scenario a robot has to determine its location in an open-plan university office based on the current image from its onboard camera and a set of pre-learned appearance models of several locations. Since the appearance of these locations changes over time, it is beneficial to utilise appearance models that explicitly represent the appearance variations [37], [5], [8], [6]. This experiment compares the impact of different temporal models, which predict the visibility of environmental features at these locations, on the robustness of robot localisation. To gather data about the changes in feature visibility, a SCITOS-G5 robot visited eight different locations of the university office every 10 minutes for one week, collecting a training dataset with more than 8000 images. After one week, the robot visited the same locations every 10 minutes for one day, collecting 1152 time-stamped images used for testing. The training set
images were then processed by the BRIEF method [38], which shows good robustness to appearance changes [39]. The extracted features belonging to the same locations were matched and we obtained their visibility over time, which was then processed by the temporal models evaluated. Thus, we obtained a dynamic appearance-based model of each location that can predict which features are likely to be visible at a particular time.

During testing, the robot uses these models to calculate the likelihood of the features’ visibility at each of the locations at the time it captured an image by its onboard camera (or extracted a time-stamped image from the testing set). In particular, it selects the \( n \) most likely-to-be-visible features at each location and time, matches these features to the features extracted from its onboard camera (or testing set) image, and determines the model with the most matches as its current location. The localisation error is calculated as the ratio of cases when the robot incorrectly estimated its location to the total number of images in the testing set. The dependence of the average localisation error on the particular temporal model and number of features \( n \) used for localisation is shown in Figure 3. The results indicate that the localisation robustness of the methods that take into account the rhythmic nature of the appearance changes outperform the Mean method, which relies on the most stable image features. Moreover, the methods that model these cyclic changes in a continuous manner perform better than the Hist method which models different times of the day in separate, as shown in Figure 3

**C. Velocity prediction**

This scenario concerns the ability of our representation to predict the velocity of a robot while navigating through a given area, which depends on how cautiously it has to move due to the presence of humans. Thus, this experiment is concerned with the ability of our method to predict a one-dimensional continuous variable (robot velocity) for a given time and location.

The velocities and times of navigation for our evaluation were obtained from a database obtained with a SCITOS-G5 mobile platform, which gathered data in an open plan research office for more than 10 weeks. Typically, the average velocity of the robot did not show much variation, but in cases it had to navigate close to workspaces and through doors, the velocity varied significantly. To evaluate the ability of our approach to predict the robot velocity, we split the dataset into an 8-weeks long training set and two testing sets of 1-week duration. As in the case of door state prediction, we calculated the mean square error of the predictions provided by our models, and compared them to find out which of the methods provide the most accurate predictions. Our results indicated that both Hypertime-based methods outperformed the other ones, as shown in Figure 4.

**D. Human presence**

Finally, we validated the proposed approach on 2-dimensional data indicating the positions of people in several corridors of the Isaac Newton Building at the University of Lincoln. Data collection was performed by a mobile robot equipped with a Velodyne 3d laser rangefinder, which was placed at a T-shaped junction so that its laser rangefinder was able to scan the three connecting corridors simultaneously. To detect and localize people in the 3d point clouds provided by the scanner, we used an efficient and reliable person detection method [40]. Since we needed to recharge the robot occasionally, we did not collect data on a 24/7 basis and recharged the robot batteries during nights, when the building is vacant and there are no people in the corridors. Thus, our dataset spanned from early mornings to late evening over several weekdays. Each day contains approximately 28000 entries, which correspond to hundreds of walks of people through the monitored corridor. To
quantitatively evaluate the model quality, we again split the gathered data into training and test sets, and learn the model from the training set only. Then, we partition the timeline of the test data into a spatio-temporal 3d grid. For each cell \( g \), we count the number of detections \( d_\text{g} \) that occurred and compare this value with the value \( p_\text{g} \) predicted by a given spatio-temporal model. To better visualise the methods’ prediction improvements, we show the reduction of the mean square error compared to the Mean model in Figure 5. To make a comparison with other models, we apply the FreMEn method on each of the grid cells independently and then predict the most likely number of events at a given time in a particular cell. Since the error is dependent on the partitioning used, we tested the method for grids of various cell sizes ranging from 5 to 20 cm and 5 to 30 minutes.

To demonstrate the model’s ability to estimate the spatio-temporal distribution over time, we let it predict the most likely occurrence of people for different times. Figure 1 and video [7] show that the predicted distributions of the people likely occurrence of people for different times. Figure 1 and video [7] show that the predicted distributions of the people likely occurrence of people for different times. Figure 1 and video [7] show that the predicted distributions of the people likely occurrence of people for different times. Figure 1 and video [7] show that the predicted distributions of the people likely occurrence of people for different times. Figure 1 and video [7] show that the predicted distributions of the people likely occurrence of people for different times. Figure 1 and video [7] show that the predicted distributions of the people likely occurrence of people for different times. Figure 1 and video [7] show that the predicted distributions of the people likely occurrence of people for different times.

VI. Conclusion

We presented a novel approach for spatio-temporal modeling for robots that are required to operate for long periods of time in changing environments. The method models the time domain in a multi-dimensional hyperspace, where each pair of dimensions represents one periodicity observed in the data. This multi-dimensional, warped time model is used to extend the state space representing a given phenomenon. By projecting the robot’s observations into this space-hypertime and clustering them, we create a continuous, spatio-temporal model (distribution) of the phenomenon observed by the robot. Knowledge of the spatio-temporal distribution is then used to predict the occurrence of a given phenomenon at a given time.

Using data collected by a mobile robot over several weeks, we show that the method can represent the spatio-temporal dynamics of binary and continuous variables, and use the representation to make predictions of the future environment states, resulting in significantly better performance than the previous state of the art.

One of the major problems we have encountered is the instability of clustering methods on the multi-dimensional hypertime space. Thus, in the future, we will evaluate the performance of different clustering algorithms on the proposed representation. Furthermore, the observed distribution is heavily influenced by the way in which the robot samples the environment, i.e. where and when it takes the measurements. Thus, we will study spatio-temporal exploration methods, which will allow a mobile robot to automatically select a location and time to obtain data useful to refine and improve the spatio-temporal model.

Finally, to allow use of the method by other researchers, we provide its baseline open source code and datasets at http://fremen.uk

REFERENCES

[1] N. Hawes et al., “The strands project: Long-term autonomy in everyday environments,” IEEE Robotics and Automation Magazine, 2016.
[2] P. Biber and T. Duckett, “Experimental analysis of sample-based maps for long-term slam,” International Journal of Robotics Research, 2009.
[3] G. D. Tipaldi et al., “Lifelong localization in changing environments,” The International Journal of Robotics Research, 2013.
[4] T. Kucner, J. Saarinen, M. Magnusson, and A. J. Lilienthal, “Conditional transition maps: Learning motion patterns in dynamic environments,” in IROS, 2013.
[5] T. Krajník, J. P. Fentanes, J. Santos, and T. Duckett, “FreMEn: Frequency map enhancement for long-term mobile robot autonomy in changing environments,” IEEE Transactions on Robotics, 2017.
[6] W. S. Churchill and P. Newman, “Experience-based navigation for long-term localisation,” IJRR, 2013.
[7] T. Krajník, “Warped hypertime representations for long-term mobile robot autonomy,” 2018. [Online]. Available: https://youtu.be/[4SW4j7DDxYE]
[8] D. M. Rosen, J. Mason, and J. J. Leonard, “Towards lifelong feature-based mapping in semi-static environments,” in International Conference on Robotics and Automation (ICRA). IEEE, May 2016, pp. 1063–1070.
[9] L. Mudrová, B. Lacerda, and N. Hawes, “An integrated control framework for long-term autonomy in mobile service robots,” in Proc. of the 7th European Conf. on Mobile Robotics (ECMR), Lincoln, United Kingdom, 2015.
[10] M. Milford and G. Wyeth, “Persistent navigation and mapping using a biologically inspired SLAM system,” The International Journal of Robotics Research, vol. 29, no. 9, pp. 1131–1153, 2010.
[11] K. Konolige and J. Bowman, “Towards lifelong visual maps,” in International Conference on Intelligent Robots and Systems, 2009.
[12] J. Pulido Fentanes, B. Lacerda, and N. Hawes, “An integrated control framework for long-term autonomy in mobile service robots,” in Proc. of the 7th European Conf. on Mobile Robotics (ECMR), Lincoln, United Kingdom, 2015.
[13] 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.
[14] T. Faellhammer, R. Ambrus, C. Burbridge, M. Zillic, J. Folkesson, N. Hawes, P. Jensfelt, and M. Vincze, “Autonomous learning of object models on a mobile robot,” Robotics and Automation Letters, 2016.
[15] D. Arbuckle, A. Howard, and M. Mataric, “Temporal occupancy grids: a method for classifying the spatio-temporal properties of the environment,” in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), vol. 1, 2002, pp. 409–414 vol.1.
[16] F. Dayoub, G. Cielniak, and T. Duckett, “Long-term experiments with partially observable environments,” in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on. IEEE, May 2016, pp. 1196–1201.
[17] 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.
[18] T. Krajník, J. Saarinen, M. Magnusson, and A. J. Lilienthal, “Conditional transition maps: Learning motion patterns in dynamic environments,” in Intelligent Robots and Systems (IROS), 2014. 2014 IEEE/RSJ International Conference on. IEEE, 2014, pp. 1832–1838.
[19] 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.
[20] 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.
[24] S. T. O’Callaghan and F. T. Ramos, “Gaussian process occupancy maps,” The International Journal of Robotics Research, vol. 31, no. 1, pp. 42–62, 2012.

[25] 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.

[26] D. E. Gustafson and W. C. Kessel, “Fuzzy clustering with a fuzzy covariance matrix,” in Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on. IEEE, 1979, pp. 761–766.

[27] J. C. Dunn, “A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters,” Journal of Cybernetics, 1973.

[28] T. Vintr, L. Pastorek, and H. Rezankova, “Autonomous robot navigation based on clustering across images,” Research and Education in Robotics-EUROBOT 2011, pp. 310–320, 2011.

[29] R. Kruse, C. Döring, and M.-J. Lesot, “Fundamentals of fuzzy clustering,” Advances in fuzzy clustering and its applications, pp. 3–30, 2007.

[30] J. Bezdek, Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers, 1981.

[31] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” in Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1, no. 14. California, USA, 1967, pp. 281–297.

[32] P. J. Groenen, U. Kaymak, and J. van Rosmalen, “Fuzzy clustering with minkowski distance functions,” Fuzzy Clustering and its Applications, Wiley, pp. 53–68, 2007.

[33] 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.

[34] A. Roy, S. K. Parui, and U. Roy, “Swgmm: a semi-wrapped gaussian mixture model for clustering of circular–linear data,” Pattern Analysis and Applications, vol. 19, no. 3, pp. 631–645, 2016.

[35] T. Krajnik, M. Hanheide, T. Vintr, K. Kusumam, and T. Duckett, “Towards automated benchmarking of robotic experiments,” in ICRA Workshop on Reproducible Research in Robotics, 2017.

[36] T. Krajnik, “The frequency map enhancement (FreMEn) project repository,” [http://fremen.uk](http://fremen.uk)

[37] P. Neubert, N. Sunderhauf, and P. Protzel, “Superpixel-based appearance change prediction for long-term navigation across seasons,” RAS, 2014.

[38] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, “BRIEF: binary robust independent elementary features,” in Proceedings of the ICCV, 2010.

[39] T. Krajnik, P. Cristóforis, K. Kusumam, P. Neubert, and T. Duckett, “Image features for visual teach-and-repeat navigation in changing environments,” Robotics and Autonomous Systems, 2016.

[40] Z. Yan, T. Duckett, N. Bellotto et al., “Online learning for human classification in 3d lidar-based tracking,” in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017.