Warped Hypertime Representations for Long-term Autonomy of Mobile Robots

Tomáš Krajník[1], Tomáš Vintr[1], Sergi Molina[2], Jaime Pulido Fentanes[2],
Grzegorz Cielniak[2], Oscar Martinez Mozos[3], George Broughton[1], Tom Duckett[2]

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 or natural processes. 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 key idea is to extend the spatial model with a set of wrapped time dimensions that represent the periodicities of the observed events. By performing clustering over this extended representation, we obtain a model that allows the prediction of probabilistic distributions of future states and events in both discrete and continuous spatial representations. We apply the proposed algorithm to several long-term datasets acquired by mobile robots 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.

Index Terms—Mapping, Learning and Adaptive Systems, Service Robots

Fig. 1. Method overview: The data points (a,t) observed over time (top, black) are first processed by frequency analysis [5] to determine a dominant periodicity T. Then, the time t is projected onto a 2d space (called hypertime) and the vectors (a,t) become $(a, cos(2\pi\,t/T), sin(2\pi\,t/T))$ (bottom, left). The projected data are then clustered (bottom, center, blue) to estimate the distribution of a over the hypertime space (green). Projection of the distribution back to the uni-dimensional time domain allows to calculate the probabilistic distribution of a for any past or future time.

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. Many authors have shown that environment models which adapt to changes improve the overall ability of mobile robots to operate over longer time periods [2]–[6]. Since long-term autonomous operation improves the chances of observing the environment changes, mobile robots gain the opportunity to learn the environment structure, but also how it changes over time [1].

Methods that explicitly represent evolution of the environment over time endow the robots with the ability to predict the future state of the world, which has a positive impact
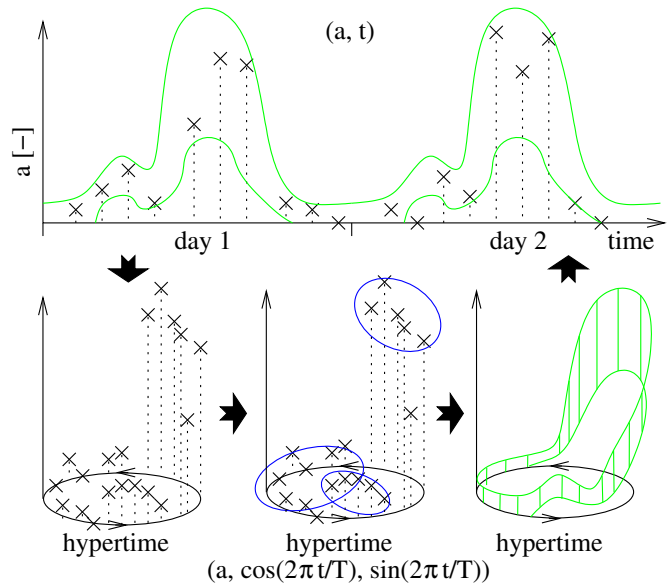
[1] Tomáš Krajník, Tomáš Vintr and George Broughton are with the Artificial Intelligence Center, Faculty of Electrical Engineering, Czech Technical University, CZ `name.surname@fel.cvut.cz`
[2] Sergi Molina, Jaime Pulido Fentanes, Grzegorz Cielniak and Tom Duckett are with Lincoln Centre for Autonomous Systems, University of Lincoln, UK
[3] Oscar Mozos is with the Technical University of Cartagena, Spain
Digital Object Identifier (DOI): see top of this page.

on the efficiency of their long-term operation. For example, models of movement patterns of objects [7] or people [4], [8] improve motion planning, and explicit models of the persistence [3], [9] or periodicity [5] of the environment states improve visual localisation. However, the latter methods are applicable to a wider range of scenarios. In particular, Frequency Map Enhancement (FreMEn) [5] can introduce dynamics into most discrete environmental representations used in robotics, and was used to improve robotic mapping [5], [10], localisation [5], [11], [12], path planning [13], robotic search [14], activity recognition [15], patrolling [16], [17], exploration [18], [19], task scheduling [20] and human-robot interaction [21], [22]. Moreover, FreMEn allows for temporal-context-based novelty and anomaly detection [5], [10], [23], and it was shown to outperform other temporal models both in prediction accuracy and computational efficiency [14]–[16], [19], [22], [23]. FreMEn is based on the assumption that some of the mid- to long-term processes that exhibit themselves through environment changes are periodic, e.g. seasonal foliage changes, day-night illumination or human

activities characterised by regular routines. By representing the dynamics of the individual environment model components in the frequency domain [5], FreMEn can efficiently store past observations, and provide accurate predictions of the future environment states. However, it was tailored to models consisting of components with binary states, e.g. occupancy grids, landmarks or topological maps.

In this paper, we present a method that also aims to explicitly model periodic changes of the environment, but it can be applied to continuous, multi-dimensional representations, and can predict the probability distribution of a given random variable at a particular time and location. The method first uses FreMEn [5] to identify (temporally) periodic patterns in the data gathered. Then, it transforms each time periodicity into a pair of dimensions that form a circle in $2d$ space and adds (concatenates) these dimensions to the vectors that represent the spatial aspects of the modelled phenomena. Finally, a generalised model is built by applying traditional techniques like clustering or expectation-maximisation over the warped time-space representation. The resulting multi-modal model represents both the structure of the space and temporal patterns of the changes or events as shown in Fig. 1. In this way, the proposed method can turn a spatial representation into a spatio-temporal one by extending it with several wrapped dimensions representing time, with each pair of temporal dimensions representing a given periodicity observed in the gathered data. We hypothesise that since the proposed model respects the spatio-temporal continuity of the modelled phenomena, it provides more accurate predictions than models which partition the space into discrete elements, or that models which neglect the temporal aspects.

We provide a description of the proposed method and experimental evidence of its capability to efficiently represent spatio-temporal data and to predict future states of the environment. Unlike the previous works [3]–[5], [9], which aim to introduce time into models that represent the environment by a discrete set of binary states, such as the 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 [24].

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], [6], [25]–[27] proposed approaches that create, refine and update world models in a continuous manner. Furthermore, Ambrus et al. [28] 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 [29].

Unlike the aforementioned works, which focused on the spatial structure and appearance aspects of the changes observed, other authors [2], [3], [5], [30], [31] focused on modelling the temporal aspects. For example, [2] and [30] 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. [31] and Rosen et al. [9] 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 [2], [3], [5], but also planning [13], [14] and exploration [18]. 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], [18], visibility of landmarks [5], traversability of edges in topological maps [13] or human presence in a particular room [14]. 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 [4], [32]. Kuczner et al. [4] 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 [32], 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, [33], [34] 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, [35] shows how to use this representation for robot motion planning in crowded environments.

O'Callaghan and Ramos [36] also argue in favour of con-

tinuous models, showing the advantages of Gaussian Mixture-based representations in terms of memory efficiency and utility for mobile robot navigation. Authors of [37] speed up building and updating of the proposed models by using an elegant combination of kernels and optimization methods. Moreover, the method is extended to perform short-term predictions of the environment state based on the history of recent observations [38].

Unlike the work of Ramos et al. [37], which is aimed primarily at modelling the spatial structure, and [33], 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 [33], [37] 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 methods, that perform predictions based on recent observations, our method can accurately predict the environment states using observations gathered days to months before the prediction. As demonstrated in [23], the predictive accuracy of FreMEn and hypertime exceeds popular, state-of-the-art methods like facebook's Prophet. Moreover, the STRANDS project [1] showed that the predictions provided have a positive impact on the efficiency of mobile robot operation in long-term scenarios.

III. METHOD

A. Motivation

Let us consider the scenarios of the STRANDS project [1], which utilised the FreMEn [5] method, and created robots that achieved four months of autonomy in human-populated environments. In the STRANDS deployments, the robots had to detect anomalies, provide an info-terminal service, and guide people to designated locations or to other people. Anomalies were not defined by hard-coded rules, but simply as events and activities occurring at unusual locations and times. Thus, the robots had to create models of the usual events and activities from their past experience and use outlier detection to identify anomalies. To provide an efficient info-terminal service, a robot has to position itself close to areas with a high level of pedestrian traffic. However, to avoid causing nuisance to people, it should arrive at these locations before they become crowded. Thus, the robot has to anticipate the occurrence of people at a given time and location based on its past experience. To guide people to the designated locations, the robot needs to be able to plan its path so that it avoids areas which might be blocked or congested. If the robot is supposed to guide someone to another person, it needs to know where that person usually is. To be able to perform the aforementioned tasks at all, the robot has to be able to localise itself despite the environment changes and to plan its path so that it reaches the desired locations in an efficient way. While predicting closed doors, congestions or localisation failures from recent observations is beneficial, long-term predictions based on past experience allow to construct plans that avoid these situations.

As already demonstrated in the STRANDS project [1], the efficiency of a mobile robot improves if it can predict the future states of the environment relevant to the service provided. However, the predictive engine of the STRANDS robots was based on the FreMEn method, which, in its original form, aims to represent independent binary states only. From a formal point of view, FreMEn represents a given phenomenon by n independent states s_i, and can estimate the conditional Bernoulli distribution $p(s_i = 1|t)$ of each state for past or future time t. In contrast, the method presented in this paper aims to estimate and predict $p(a|\mathbf{x}, t)$, where a is the predicted variable (e.g. landmark visibility, robot velocity, number of people present etc.), and $\mathbf{x}, t$ are the location and time. This not only alleviates the necessity to spatially discretise the modelled phenomenon, but also allows to predict other-than-Bernoulli distributions.

B. Core Idea

The problem of finding the conditional distribution $p(a|\mathbf{x}, t)$ is that while the modelled space $\mathbf{x} \in \mathbb{R}^m$ 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.

However, many events and changes occurring in human-populated environments are repetitive due to the nature of human habits. To represent the repetitive nature of these changes, we project the time t into a set of circles, where every circle is derived from the periodicity of change T_i detected by FreMEn in the measured phenomenon. This causes the time-dependent events with the same periodicity to be projected into the same areas of a circle that corresponds to the modelled periodicity, see Figure 1. The distribution of observations projected into this warped space (i.e. space with 'wrapped' or 'curled' time dimensions) can be then estimated using standard statistical and machine learning tools. To predict the probabilistic distribution of a given variable for a given time t, we simply project t to the circular space and perform the prediction there.

The projection of linear time onto a set of wrapped circles reflects the fact that human activities in the mornings of different days are more similar than human activities in the mornings and afternoons of the same day although the same day afternoon and morning is temporally closer than mornings of two different days. Furthermore, the projection reflects the fact that a given phenomenon does not change abruptly during midnight although 23:59 and 0:01 appear to be distant. An illustrative example of the method, which estimates $p(a|t)$ through projection into $p(a|cos(2\pi t/T), sin(2\pi t/T))$ is in Figure 1. For the sake of simplicity, this example uses only one periodicity T and the spatial domain is neglected, i.e. $\mathbf{x} \in \mathbb{R}^0$.

C. Method Overview

Let us assume that a robot gathered a set containing l measurements of a given phenomenon, obtaining tuples $(a_i, \mathbf{x}_i, t_i)$,

where $i \in \{1 \dots l\}$, the vector $\mathbf{x}_i$ 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 $(\mathbf{x}_i, t_i)$.

Our method aims to find a $p(a|\mathbf{x}, t)$, which would represent the conditional probability density function of the variable a given the position $\mathbf{x}$ and time t. The proposed method is composed of five stages: 1) initialization; 2) spatio-temporal clustering; 3) model error estimation; 4) identification of periodicities; 5) hypertime space extension. After initialization, our method repeatedly performs steps 2-4 as long as the model error, estimated by Eq. (6) during step 3, decreases.

D. Step I: Initialization

To initialize the algorithm, we first store all measurements $(a_i, \mathbf{x}_i)$ in $^h\mathbf{x}_i$, where h corresponds to the number of known periodicities. Since during initialisation, the number of periodicities is unknown, we set h to 0. In the *spatio-temporal clustering* stage, we cluster the vectors $(^h\mathbf{x}_i)$, obtaining a Gaussian mixture model, which represents the spatio-temporal distribution of the given phenomenon and allows to calculate conditional probability function $p_h(a|\mathbf{x}, t)$. In the *model error estimation*, we calculate the mean μ_i of $p(a|\mathbf{x}_i, t_i)$ for all training samples. Then we calculate the time series $^h\epsilon(t_i)$ as $^h\epsilon(t_i) = \mu_i - a_i$ and its mean squared value E_h, see (6) in Section III-F. Then, during the *identification of periodicities*, we use the FreMEn [5] method to 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 $^h\mathbf{x}_i$ by 2 dimensions representing a given periodicity of the temporal domain, i.e.

$$^{h+1}\mathbf{x}_i \leftarrow \left(^h\mathbf{x}_i, \cos(2\pi\, t_i/T_{h+1}), \sin(2\pi\, t_i/T_{h+1})\right). \quad (1)$$

Then, we increment h by one and repeat the steps of *spatio-temporal clustering* and *model error estimation* on the now extended vectors $^h\mathbf{x}_i$, 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 $^h\mathbf{x}_i$ 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 $p_{h-1}(a, \mathbf{x}, t)$ from previous iteration as $p(a, \mathbf{x}, t)$ and terminate the method.

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

E. Step II: Spatio-Temporal Clustering

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

$$p(a, \mathbf{x}, t) = \gamma \sum_{j=1}^{n} w_j\, u_j(a, \mathbf{x}, {}^h\mathbf{t}), \quad (2)$$

where $u_j(a, \mathbf{x}, {}^h\mathbf{t})$ is a multivariate Gaussian function of the jth cluster, w_j is the cluster weight, $^h\mathbf{t} = \left(\cos(2\pi\frac{t}{T_1}), sin(2\pi\frac{t}{T_1}), \dots, \cos(2\pi\frac{t}{T_h}), \sin(2\pi\frac{t}{T_h})\right)$ is the projection of time in the hypertime space and γ is a scaling constant, specified by (3) in Section III-F.

F. Step III: 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 γ in such a way that the mean value of a_i calculated from the model (2) over the training set vectors $(\mathbf{x_i}, t_i)$ is equal to the average value of a_i on the training set:

$$\gamma = \frac{\sum_{i=1}^{l} a_i}{\sum_{i=1}^{l} a_i \sum_{j=1}^{n} w_j\, u_j(a_i, \mathbf{x}_i, {}^h\mathbf{t_i})} \quad (3)$$

After calculating the scaling factor γ, we compute an estimate of a_i at each training set point defined by location $\mathbf{x}_i$ and time t_i by calculating the mean μ_i:

$$\mu_i = \int a\, p(a, \mathbf{x}_i, t_i)\, da \quad (4)$$

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

$$^h\epsilon(t_i) = \mu_i - a_i. \quad (5)$$

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

$$E_h = \sum_{i=1}^{l} {}^h\epsilon^2(t_i) = \sum_{i=1}^{l} (\mu_i - a_i)^2. \quad (6)$$

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}, we store the current model, represented by Equation (2), and we perform another iteration of the method. If $E_h > E_{h-1}$, we terminate the algorithm. Then, the last stored model (Eq. 2), can be used to predict conditional probabilistic distribution $p(a|\mathbf{x}, t)$. This is calculated in a numerical manner, i.e. we fix $\mathbf{x}$ and/or t at the values we want to perform the prediction for, then we calculate $p(a_c, \mathbf{x}, t)$ for values of $a_c \in \mathcal{A}$ which represent the domain of a, and finally we normalise the results to ensure that $\sum_{a_c \in \mathcal{A}} p(a_c|\mathbf{x}, t) = 1$.

G. Steps IV and V: 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 $^h\epsilon(t_i)$ by the FreMEn method [5], which, unlike traditional Discrete or Fast Fourier transforms, is suitable for finding periodicities in non-uniform and sparse data. In particular, we calculate the most prominent periodicity T_h in the error time-series as:

$$T_{h+1} = arg \max_{T_k} \sum_{i=1}^{l} |(^h\epsilon(t_i) -^h \hat{\epsilon}) \, e^{-j \, 2 \, \pi \, t_i / T_k}|, \quad (7)$$

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

Thus, at the start of our method, each vector ^0x_i contains only the spatial information, i.e. $^0x_i = (a_i, \mathbf{x_i})$, but at the end, the vector contains $2h$ additional dimensions modelling the periodicities observed in the training data.

H. Clustering implementation

While the hypertime extension, error estimation, and periodicity estimation steps of the method are quite straightforward, deterministic and computationally inexpensive, the way to build the model of the probability distribution over the hypertime space is key to the method's predictive accurracy. The main issue of the hypertime space is its sparsity, because the time, which is linear and one-dimensional, is projected onto a multi-dimensional hyper-torus. With the growing number of modelled periodicities (and thus, the number of temporal dimensions), many algorithms that we tested exhibited numerical instability. Thus, we dedicated a significant effort into testing various clustering methods, their initialisations and metrics [39].

In our experiments, we utilise two clustering methods, which, in our previous experiments, provided the most satisfactory results: the HyperTime Expectation Maximisation (*HyT-EM*) and HyperTime 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 method name as *HyT-EM_k*, where k is the number of clusters used during the experiments (Section IV). To deal with the occasional numerical instabilities of the OpenCV's EM implementation, the *HyT-EM* method performs eigenvalue analysis of the model's covariance matrices and if necessary, it restarts the EM with different initial positions of the clusters.

The *HyT-KM* method was designed to deal with the aforementioned issues, making it more suitable for clustering over the hypertime space, where the training data become sparser as the number of temporal dimensions increases. *HyT-KM* first initialises the cluster centres using k-means based clustering. Inspired by [33], [35], [40], we use the mixtures of cosine

distance for hypertime and the Euclidean metrics for other variables. After initialisation, it calculates the covariace matrices of the clusters and then it proceeds with the Expectation Maximisation procedure, while using the cosine distance for temporal dimensions. In contrast to *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. It builds models with n and $n+1$ clusters and calculates the sum of amplitutes $T_\Sigma(n)$ and $T_\Sigma(n+1)$ of the frequency spectrum of the error $^h\epsilon(t_i)$

$$T_\Sigma(n) = \sum_{k=1}^{K} \sum_{i=1}^{l} |(^h\epsilon(t_i) -^h \hat{\epsilon}) \, e^{-j \, 2 \, \pi \, t_i / T_k}|. \quad (8)$$

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

IV. 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 relate to a real deployment of a mobile robot in human populated environment, see Section III-C and [1], 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_n*, 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 III-H. The experimental evaluation is performed by an automated system [41], 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 in terms of their prediction error. The results of the statistical evaluations are shown on the right sides of the Figures 2-5, where an arrow from A to B indicates that method A achieved statistically significantly less erroneous predictions than method B. To enable the reproducibility of the results, the evaluation system, source codes and datasets are available online [42].

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 ϵ of the various temporal models' predictions $p(t)$ to the ground truth $s(t)$ as $\epsilon = \sum_T (p(t) - s(t))^2 / |s(t)|$. The results shown in Fig. 2 indicate
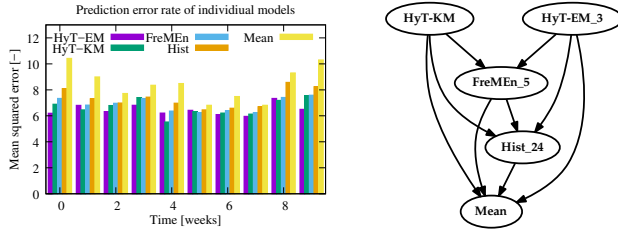

Fig. 2. Door state prediction error. The left figure shows the MSE for the training (week 0) and testing (weeks 1-9) datasets. An arrow from model A to model B in the right figure indicates that A's prediction error is statistically significantly lower than prediction error of model B.

that both hypertime-based models outperformed the other ones, including FreMEn [5]. Similarly to the results of [5], both methods indicated that that the best predictions are achieved by modeling three most prominent periodicies. In this experiment, these corresponded to one day, four hours and one week.

The reason behind the superiority of the hypertime-based methods is that FreMen approximates the training data by only one (sinus) function per observed periodicity. This causes difficulties when modeling states, which are influenced by several processes with the same periodicity or when representing short duration, regular events. The hypertime-based methods combine the advantages of mixture-based representations, which can approximate arbitrarily-shaped, multimodal functions, with the ability of FreMEn to represent their natural periodicities. For an illustrative example of the advantages of GMM and FreMEn models, leading to their combination in this work, please see our earlier work [14].

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 [5], [6], [9], [43]. 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 [44], which shows good robustness to appearance changes [45]. 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
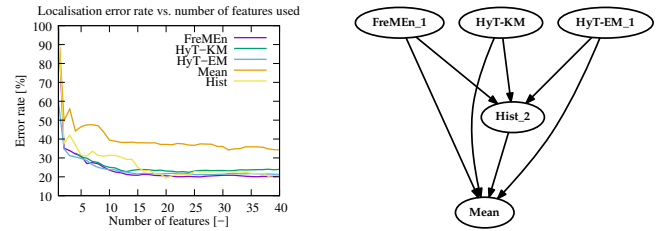

Fig. 3. Temporal model performance for feature-based topological localisation. The left figure shows the dependence of localisation error rate on the number of features predicted by a given temporal model. An arrow from A to B in the right indicates that A's localisation error rate is statistically significantly lower than localisation error rate of model B.

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. Since the robot was used for other purposes, the data were not gathered in a regular way, but

contain long hiatuses, see [13]. 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
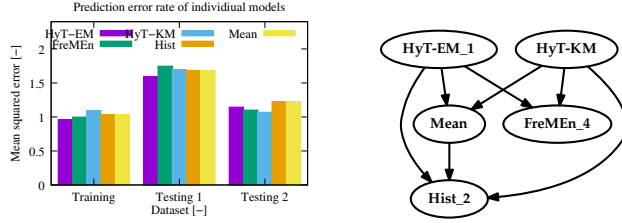

Fig. 4. Navigation velocity reconstruction and prediction errors. The left figure shows the MSE for the training and testing sets. An arrow from A to B in the right indicates that A's prediction error is statistically significantly lower than velocity prediction error of model B.

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. While the HyT-KM reconstruction error (i.e. error calculated on the training data) is higher, its prediction error is lower compared to *Mean*, *FreMEn* and *Hist* methods, see Figure 4. Since the velocity of the robot is always between 0 and 1, we applied FreMEn directly to the velocity values $v \in\, <0, 1>$ and considered the FreMEn prediction for a particular time a velocity. The same scheme was used to predict the robot velocity in [1].

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 range-finder 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 [46]. 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. Similarly to the previous case, the robot was often used for other purposes, resulting in large gaps in the data collected. 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 by 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_g that occurred and compare this value with the value p_g predicted by a given spatio-temporal model. To better visualise
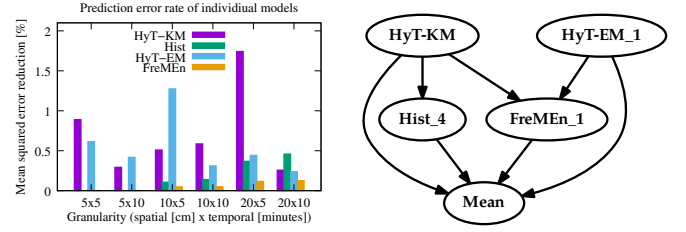

Fig. 5. Human presence prediction results. The left figure shows how the mean squared error reduced for a particular model granularity compared to the *Mean* model. An arrow from A to B in the right indicates that A's prediction error is statistically significantly lower than velocity prediction error of model B.

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. Since the number of people passing through a cell follows a Poisson distribution, we employed FreMEn modification proposed in [22], which allows to process non-binary values. However, the number of passing people exceeds one only in rare occations, and one would obtain almost the same results by using the original FreMEn [5].

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 and composed a video [47] showing how the predicted distributions of the people depend on time.

V. 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 probabilistic distribution of a given phenomenon at a given time and location.

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.

The main contribution of the article does not rest in the clustering methods, but in the alternative representation of the time domain, which, unlike the linear time, respects the

repetitiveness of events and changes imposed by naturally-occurring cycles. Moreover, we outline a formulation of the problem, which allows to apply the aforementioned method to most environmental models used in the robotic domain. We believe that this human-oriented circular representation of time will allow service robots to synchronise their activities with the habits of people they are supposed to serve.

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 considering models capable of treating spatial and temporal domains in separate or employing only a subset of the detected periodicities [48]. 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. Moreover, we will extend the spatio-temporal modeling approach towards its ability to take into account sensor noise in way similar to standard Bayesian update applied to states with fixed probabilistic values.

To allow use of the method by other researchers, we provide its baseline open source code and datasets at chronorobotics.tk and fremen.uk.

REFERENCES

[1] N. Hawes *et al.*, "The STRANDS project: Long-term autonomy in everyday environments," *IEEE Robotics & Automation Mag.*, 2017.
[2] P. Biber and T. Duckett, "Experimental analysis of sample-based maps for long-term SLAM," *Int. 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 *et al.*, "Conditional transition maps: Learning motion patterns in dynamic environments," in *IROS*, 2013.
[5] T. Krajník *et al.*, "Fremen: Frequency map enhancement for long-term mobile robot autonomy in changing environments," *T-RO*, 2017.
[6] W. S. Churchill and P. Newman, "Experience-based navigation for long-term localisation," *IJRR*, 2013.
[7] N. Bore *et al.*, "Efficient retrieval of arbitrary objects from long-term robot observations," *Robotics and Autonomous Systems*, 2017.
[8] S. Molina *et al.*, "Go with the flow: Exploration and mapping of pedestrian flow patterns from partial observations," in *ICRA*, 2019.
[9] D. M. Rosen, J. Mason, and J. J. Leonard, "Towards lifelong feature-based mapping in semi-static environments," in *ICRA*, 2016.
[10] T. Krajník, J. P. Fentanes, G. Cielniak, C. Dondrup, and T. Duckett, "Spectral analysis for long-term robotic mapping," in *ICRA*, 2014.
[11] T. Krajník *et al.*, "Long-term topological localization for service robots in dynamic environments using spectral maps," in *IROS*, 2014.
[12] J. P. Fentanes *et al.*, "Persistent localization and life-long mapping in changing environments using the frequency map enhancement," in *Int. Conference on Intelligent Robots and Systems (IROS)*, 2016.
[13] J. P.Fentanes *et al.*, "Now or later? predicting and maximising success of navigation actions from long-term experience," in *ICRA*, 2015.
[14] T. Krajník *et al.*, "Where's waldo at time t? using spatio-temporal models for mobile robot search," in *ICRA*, 2015.
[15] C. Coppola, T. Krajník, N. Bellotto, and T. Duckett, "Learning temporal context for activity recognition," in *ECAI*, 2016.
[16] T. Krajník, J. M. Santos, and T. Duckett, "Life-long spatio-temporal exploration of dynamic environments," in *ECMR*, 2015.
[17] L. Halodová *et al.*, "Adaptive image processing methods for outdoor autonomous vehicles," in *MESAS*, 2019.
[18] J. M. Santos *et al.*, "Lifelong information-driven exploration to complete and refine 4d spatio-temporal maps," in *RA-L*, 2016.
[19] J. M. Santos, T. Krajník, and T. Duckett, "Spatio-temporal exploration strategies for long-term autonomy of mobile robots," *Robotics and Autonomous Systems*, 2016.
[20] M. Kulich *et al.*, "To explore or to exploit? learning humans behaviour to maximize interactions with them," in *MESAS*, 2016.
[21] 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.
[22] F. Jovan *et al.*, "A Poisson-Spectral Model for Modelling the Spatio-Temporal Patterns in Human Data," in *IROS*, 2016.
[23] T. Vintr *et al.*, "Spatiotemporal models of human activity for robotic patrolling," in *MESAS*, 2019.
[24] L. Mudrová *et al.*, "An integrated control framework for long-term autonomy in mobile service robots," in *ECMR*, 2015.
[25] M. Milford and G. Wyeth, "Persistent navigation and mapping using a biologically inspired SLAM system," *IJRR*, 2010.
[26] K. Konolige and J. Bowman, "Towards lifelong visual maps," in *International Conference on Intelligent Robots and Systems*, 2009.
[27] S. Hochdorfer and C. Schlegel, "Towards a robust visual SLAM approach," in *International Conference on Advanced Robotics*, 2009.
[28] R. Ambrus *et al.*, "Meta-rooms: Building and maintaining long term spatial models in a dynamic world," in *IROS*, 2014.
[29] T. Faeulhammer *et al.*, "Autonomous learning of object models on a mobile robot," *Robotics and Automation Letters*, 2016.
[30] D. Arbuckle, A. Howard, and M. Mataric, "Temporal occupancy grids: a method for classifying the spatio-temporal properties of the environment," in *IROS*, 2002.
[31] 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.
[32] Z. Wang, R. Ambrus, P. Jensfelt, and J. Folkesson, "Modeling motion patterns of dynamic objects by IOHMM," in *IROS*, 2014.
[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 RAL*, 2017.
[34] R. Senanayake and F. Ramos, "Directional grid maps: modeling multi-modal angular uncertainty in dynamic environments," in *IROS*, 2018.
[35] L. Palmieri, T. Kucner, M. Magnusson, A. Lilienthal, and K. Arras, "Kinodynamic motion planning on Gaussian mixture fields," in *ICRA*, 2017.
[36] S. T. OCallaghan and F. T. Ramos, "Gaussian process occupancy maps," *IJRR*, vol. 31, no. 1, pp. 42–62, 2012.
[37] 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.
[38] R. Senanayake, L. Ott, S. O'Callaghan, and F. T. Ramos, "Spatio-temporal hilbert maps for continuous occupancy representation in dynamic environments," in *NIPS*, 2016, pp. 3925–3933.
[39] R. Kruse, C. Döring, and M.-J. Lesot, "Fundamentals of fuzzy clustering," *Advances in fuzzy clustering and its applications*, pp. 3–30, 2007.
[40] 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.
[41] 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.
[42] T. Krajník, "The frequency map enhancement (FreMEn) project repository," http://fremen.uk.
[43] P. Neubert, N. Sünderhauf, and P. Protzel, "Superpixel-based appearance change prediction for long-term navigation across seasons," *RAS*, 2014.
[44] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: binary robust independent elementary features," in *Proceedings of the ICCV*, 2010.
[45] T. Krajník, 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.
[46] Z. Yan, T. Duckett, N. Bellotto *et al.*, "Online learning for human classification in 3d lidar-based tracking," in *IROS*. IEEE, 2017.
[47] T. Krajník, "Warped hypertime representations for long-term mobile robot autonomy," 2018. [Online]. Available: https://youtu.be/4SW4j7DDxYE
[48] N. Cressie and H. Hsin-Cheng, "Classes of nonseparable, spatio-temporal stationary covariance functions," *Journal of the American Statistical Association*, 1999.