Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Computer Science

Maps of Dynamics for Social-Aware Navigation of Autonomous Mobile Robots

Disertation thesis

Ing. Tomáš Vintr

Prague, July 2023

Ph.D. programme: Electrical Engineering and Information Technology
Branch of study: Information Science and Computer Engineering
Supervisor: Doc. Ing. Tomáš Krajník, Ph.D.

Thesis Supervisor:
Doc. Ing. Tomas Krajn k, Ph.D.
Department of Computer Science
Faculty of Electrical Engineering
Czech Technical University in Prague
Technicka 2
160 00 Prague 6
Czech Republic

Supervisor Specialist:
Prof. Ing. Jan Faigl, Ph.D.
Department of Computer Science
Faculty of Electrical Engineering
Czech Technical University in Prague
Technicka 2
160 00 Prague 6
Czech Republic

Declaration

I hereby declare I have written this doctoral thesis independently and quoted all the sources of information used in accordance with methodological instructions on ethical principles for writing an academic thesis. Moreover, I state that this thesis has neither been submitted nor accepted for any other degree.

In Prague, July 2023

....
Ing. Tomas Vintr

Abstract

Robots are expected to help people with many different tasks. We hypothesise that apart from the ability to solve the task, service robots need also integrate into society to provide a service in an acceptable way. Human society has rules that follow working routines, traditions, and derived habits. Breaking the rules by any entity, including autonomous robots, spoils the psychosocial atmosphere and leads to the consequent exclusion of a maladjusted entity - in the case of the robot, refusal of its assistance.

In this work, we propose a way how an autonomous service robot could, at least, blend into human society - applying spatio-temporal maps allowing for human-induced dynamics forecasting into its planning module. We defined and evaluated spatio-temporal models suitable for autonomous service robots in various scenarios. We proposed a methodology for defining a miscellaneous collection of criteria for evaluating and comparing diverse spatio-temporal maps from the point of usability in specific robotic tasks. We derived a specific criterion for evaluating spatio-temporal maps for human-aware trajectory scheduling and planning. Using this criterion, we compared an exhaustive collection of simplified approaches to spatio-temporal mapping, providing a reader with an insight into their differences, and the current state-of-the-art methods, including the proposed one, providing a complex assessment of their strengths and weaknesses. We also evaluated the impact of integrating the proposed spatio-temporal map into the planning module in a field robotic experiment that analysed reactions of uninformed human subjects.

The results of the field experiment showed a radical difference in reactions towards robot that was and was not able to forecast human flows. A robot not using a spatio-temporal map annoyed some people, and one human got angry, while the robot able to blend into the human flows solved its task unnoticed. The proposed spatio-temporal modelling method successfully competed with its competitors. It proved its qualities when integrated into a public application. It provided forecasts covering country-wide area from very sparse data.

Keywords: spatio-temporal forecasting, long-term autonomy, human-aware task scheduling, comparing spatio-temporal maps, chronorobotics, maps of dynamics

Anotace

Od robotu se ocekava, ze budou pomahat lidem s mnoha ruznymi ukoly. Predpokladame, ze krome schopnosti resit ukol se servisn roboti potrebuj take zaclenit do spolecnosti, aby poskytovali sluzbu prijatelnym zpusobem. Lidska spolecnost ma pravidla, ktera se rd pracovnmi rutinami, tradicemi a odvozenymi zvyky. Porusen pravidel jakoukoliv entitou, vcetne autonomnch robotu, kaz psychosocialn atmosferu a vede k naslednemu vyloucen neprizpusobive entity { v prpade robota k odmtnut jeho pomoci.

V teto praci navrhujeme zpusob, jak by se autonomn obsluzny robot mohl alespon zaclenit do lidske spolecnosti { aplikovanm casoprostorovych map umoznujcch predpovedi dynamiky vyvolane clovekem do jeho planovacho modulu. De novali a vyhodnocovali jsme casoprostorove modely vhodne pro autonomn servisn roboty v ruznych scenarch. Navrhli jsme metodiku pro de - novan ruznorodeho souboru kriteri pro hodnocen a porovnavan ruznych casoprostorovych map z hlediska pouzitelnosti v konkretnch robotickych ulohach. Odvodili jsme speci cke kriterium pro vyhodnocen casoprostorovych map pro rozvrzen a planovan trajektori s ohledem na lidi. Pomoc tohoto kriteria jsme porovnali vycerpavajc sbrku zjednodusenych prstupu k casoprostorovemu mapovan, ktere ctenari poskytly nahled na jejich odlisnosti, a soucasnych nejmodernejsch metod, vcetne te navrhovane, poskytujc komplexn posouzen jejich silnych a slabych stranek. Hodnotili jsme take dopad integrace navrzene casoprostorove mapy do planovacho modulu v polnm robotickem experimentu, ktery analyzoval reakce neinformovanych lidskych subjektu.

Vysledky polnho experimentu ukazaly radikaln rozdl v reakcch na robota, ktery byl a nebyl schopen predpovdat lidske toky. Robot, ktery nepouzval casoprostorovou mapu, nektere lidi nastval a jednoho cloveka rozzlobil, zatmco robot schopny nasledovat lidske toky svuj ukol nepozorovane vyresil. Navrzena metoda casoprostoroveho modelovan uspesne konkurovala svym konkurentum. Sve kvality prokazala pri integraci do verejne aplikace. Poskytla predpovedi pokryvajc celou zemi z velmi rdkych udaju.

Klcova Slova: casoprostorove progrnozovan, dlouhodoba autonomie, planovan ukolu s ohledem na cloveka, porovnavan casoprostorovych map, chronorobotika, mapy dynamiky

Acknowledgements

This dissertation thesis was supported by:

- Czech Ministry of Education - \3L4AR", project ID 8J23FR023,
- EU H2020 FET - \RoboRoyale", project ID 964492,
- Czech Science Foundation - \ToltaTempo", project ID 20-27034J,
- Ministry of Industry and Trade - \System for Social Temperance Based on Time-Space Models", project ID IDENT189,
- OP VVV - \Research Center for Informatics", project ID CZ.02.101/0.0/0.0/16_019/0000765,
- Czech Ministry of Education - \Autonomous vehicles", project ID FR-8J18FR018,
- and Czech Science Foundation - Spatio Temporal Representations for Life-Long Mobile Robot Navigation, project ID GJ17-27006Y.

List of Tables

List of Figures

Contents

Part Ⅰ

Introduction

Starting Point

Thanks to computer hardware advances in recent years and the consequent advances in arti cial intelligence, society expects robots to help people with many di erent tasks. They are supposed to work in changing conditions for long periods and replace humans in repetitive tasks and tasks where humans fail due to their emotional nature, slow reactions or low dexterity. They are expected to excel in tasks requiring long-term focus for rapid and precise reactions to rare and unexpected situations. Due to the massive data ow and progress in the databases, autonomous systems are now understood as knowledge providers that can help inexperienced humans reason in unusual situations or choose the best solution before more experienced humans arrive.

Robots can operate in controlled and precisely mapped environments. The current aim is to make them work in an unknown and unaccustomed environment for a long time. Such a general environment evinces various irregularities - changes over time due to natural processes or human activity. These changes in the environment lead to the gradual worsening of the static maps, and it is not possible only to change the obsolete static map with the new one [1]. The explicit representation of these changes over time prolongs the time an autonomous robot can carry out its duties [2]{[8] and rises the quality and performance in long-term tasks [9]{[13].

This doctoral thesis summarises the latest advances in modelling the dynamics of the robot operational area. It describes and discusses methods proposed during my study. Those methods address issues that robots face when operating in dynamic, human-populated environments. An integral part of this work was a collection of long-term datasets used to evaluate developed methods.

Original Research Objectives

The original goal of my dissertation thesis was to create a general tool to model spatio-temporal phenomena useful for robots operating in large, human-populated areas for long periods.

We expected that a human-populated environment would typically incorporate changes in structure derived from open or closed doors, the number of people present, visual appearance due to diurnal rhythms, dominant directions of pedestrian flow, and the crowdedness of different areas. It was known that predicting such phenomena and adapting a robot to the actual state leads to better performance in accomplishing its tasks [9]. However, different predictions of different variables in different scenarios lead to different machine-learning-based approaches. Therefore we wanted to define the universal method applicable to:

- Binary Classification (robotic maps often consist of components with binary states),
- (Event) Frequency Analysis (for densities and flows estimation),
- Multiclass Classification (for regular activity recognition),
- Regression (for forecasting a value of a quantity, like speed and crowdedness, at a given location),
- and Novelty Detection (to deal with rare events and changes of dynamics)

for long-term spatio-temporal mapping. To achieve the goals, we decided to extend the existing collection of datasets:

- long-term person presence locations [14],
- and long-term open/closed door detection [15],

by collecting new ones:

- long-term spatial human detection on a 24/7 basis,
- navigation maps in changing outdoor environment,
- and long-term spatio-directional human detections

for a complex evaluation of the method. Finally, we wanted to deploy the method on a real robot and prove its usability in a field experiment.

Text Organisation

The body of the text is divided into three chronologically ordered parts. The first part concerns modelling mid-to-long-term spatio-temporal dynamics of human-populated, office-like environments. Described methods proceed from Frequency Map Enhancement [5], expand an idea of its continuous extension, Hypertime [16], and address technical details of Hypertime implementation.

The main issue during the development of these methods was comparing and evaluating the predictive abilities of different spatio-temporal methods and maps. The second part concerns a methodology for the quality of spatio-temporal maps comparison regarding expected robotic tasks [17]. This part includes state-of-the-art methods comparison and provides a complex analysis of their differences, strengths and weaknesses. It also provides an overview of different methods for long-term spatio-temporal modelling and mapping from various fields and compares simplified approaches derived from them.

Previous and actual data collection is covered in the third part. The data collection is generally costly and time-consuming. At the same time, the success, especially in collecting month-to-year-long datasets, is uncertain. However, the methods developed on existing datasets are prone to overfitting, usually in unexpected and unnoticeable ways. This part describes our expectations on the new resulting dataset and technology used for collection until now.

Almost the whole text in this thesis is a copy of texts from reviewed papers where I was the first author. The proposed ideas result from discussion and cooperation with co-authors of original works. All the authors edited those texts, and (usually unknown) reviewers gave the reader the final impression. The first part is a collection of texts from [18]{[22], the second part consists of text from [17], [23], and the third part is mainly a copy of [24]. Two paragraphs from [22] were also used in the Conclusion.

Part II

Long-Term Spatio-Temporal Mapping for Human-Aware Autonomous Robotics

1 Outline

The rise of autonomous mobile robotics is intertwined with the development of their ability to understand the surrounding world. More recent works have shown that the maps and models representing the environment bene t from including temporal dynamics and the capability of these models to predict future environment states [25]. Such an extension has posed the question of how best to explore, capture, represent, and exploit dynamics and discover periodic patterns and rare cyclical events while ignoring trends with negligible mid-term impact. The topic is complicated because while these robots are autonomous, they operate in the real world and are subject to real-world technical limitations. For example, downtime for battery charging can create gaps and nonuniformities in observations. Nevertheless, it has been shown that modelling and consequent predictions of changes to estimate the state of maps in the future give robots the ability to operate in their environment for a longer time with better task e ciency than when using static maps [2], [3], [6].

One of the most common constraints for autonomous robots modelling the dynamics of their environments is computational e ciency due to power and time limitations. As a result, one of the most successful ways to model these changes is to use a Fourier transform approach (or Fourier transform derived), denoted Frequency Map Enhancement (FreMEn), for an implementation applied to predictive robotic maps [5]. This is a logical choice, given that the primary drivers of changes in the natural environments are periodic, whereas trends seem unimportant in the midterm.

FreMEn introduced frequency analysis into the spatio-temporal mapping and hypothesised that the periodicities of human habits (sleep, going to work) and human-induced dynamics in an environment are dominant compared to progress or evolution in the horizon of months and years. However, FreMEn is a spatially discrete model. Its main issue is the spatial independence of the neighbouring cells. Modelling spatial dependencies in the dynamic environment and problems related to the choice of the model spatial resolution [26] led to an idea of continuous spatio-temporal representation of the environment. Although the continuous models of the environment changes are computationally demanding, they outperform the discrete ones in terms of storage memory e ciency [27], which is a necessary requirement for the ability to store large spatio-temporal maps.

2 Related Work

Many robotic methods assume that environmental uncertainty comes primarily from the imperfection of the sensory input [28] and ignore the fact that the environment itself can change over time [29]. Some methods consider environmental changes that significantly influence the deployment of a robot. For example, illumination changes during the day can be mitigated by filtering the received data in the photometric domain [30], [31] or by utilising environment structure [32]. Some changes in the scenes can be modelled using Bayesian probability [33], and seasonal changes can be suppressed by sophisticated forecasting of visual appearance [34], [35].

Philosophically different approaches to environmental changes emerged from the idea that localisation and mapping methods used by autonomous robots need not focus only on accuracy but, more importantly, on flexibility, robustness and adaptability [1], [6]. The idea evolved during the STRANDS project [9], focusing on the long-term deployment of autonomous robots in natural, human-populated areas, and then expanded during the project STROLL [36]. The STRANDS consortium argued that environmental changes should not be understood as an inevitable source of the gradual degradation of well-established static models (occupancy grids, visual landmarks, edges in topological maps). Instead, the natural dynamics of the environment can be learned and exploited by long-term running autonomous systems. The STRANDS team pointed out that a large part of the dynamics relates to natural cycles, and they determined that the uncertainty in the models can be represented as a function of time. Through the identification of these routines, autonomous robots can learn from and adapt to the changes rather than trying to neglect or suppress them.

2.1 Non-FreMEn Approaches to Spatio-Temporal Mapping

The most usual way to model periodical events in the spatio-temporal maps is to include a priori known periods derived from the authors' expertise in the model's architecture. The straightforward way is to create seasonal windows that include time-speci c maps [26], [37]{[41]. Another approach is based on kernel warping. The map includes a continuous Gaussian mixture model with prede ned periodical kernels [42]{[46]. There also exist approaches of manual preprocessing the data to nd dominant periods using auto-correlation [47]{[49] and then applying the found periods into the architecture of the spatio-temporal maps.

Automatic estimation of periods in the data was targeted in [50] by applying Gaussian processes to the data of continuous media ow. The model included a covariance matrix constituted by separately calculated spatial and temporal components. Temporal components included temporal decay and periodical components estimated from the data iteratively using frequency analysis. The same group then expanded the idea by proposing Hilbert maps [51], their incremental update [52], Fourier Feature Approximations for Periodic Kernels [53] and its multidimensional variant [54]. However, the mathematical and computational complexity limits its usage in autonomous robot systems.

2.2 Fourier Transform in Long-Term Autonomy

In [55], the authors apply the Fast Fourier Transform (FFT) to the past observed binary states of an occupancy grid and de ne the occupancy states by a periodical function of time. Spectral Mapping acquired the ability to predict the states in the map. Compared to the classical occupancy grid, Spectral Mapping lowered the prediction error by 60%. They optimised the occupancy map by combining Spectral Mapping with an octree-based spatial model [56] into a 4-dimensional model of the environment, FROctomap [57], and integrated that within the Robotic Operating System (ROS). Spectral Mapping was also applied to the topological localisation task [15]. The robot visited 8 different places every 10 minute for one week and captured the visual appearance of its surroundings at every place. The images were processed into image features, and the FFT then modelled the visibility of those features. The robot created a map of features specific to a place and time. The experiments proved the ability of the robot to localise itself using the feature map even after 3 months. The authors found a correlation between the prominence of periods used in the model and the persistence of the map.

The Spectral Mapping method proved that temporal prediction gives autonomous robots an edge, but inherent FFT principles require a stable observation rate without interruption. It is hard to attain such observations outside a laboratory, especially with autonomous mobile robots. Additionally, Spectral Mapping cannot be updated by adding new data from later observations, limiting long-term deployments. Those problems were addressed by Frequency Map Enhancement [5], [58], which is based on the definition of the Non-Uniform Discrete Fourier Transform (NUDFT) [59] rather than FFT and allows building and updating models with irregular data.

FreMEn was applied to the edges of a topological map [60] as a part of the Time-Indexed Navigation Markov Decision Processes. During the experiment, the robot gathered data over 2 months and proved a significant improvement in planning the optimal time to navigate through the environment. To prove the dominance of the proposed approach in planning tasks, FreMEn was applied to a robotic search task [61]. The robot modelled the presence of people at different places in 3 different environments. During the experiment, it planned a path through the environment, intending to find a person as quickly as possible. Compared to the strategy based on the static map, which did not consider human habits, the strategies exploiting the model of usual human behaviours decreased the search time by 25% and the number of visited places by 33%. Aiming to model the number of people in certain areas instead of binary states or the probability, the authors of [62] redefine FreMEn, which resulted in an iterative approach denoted as Addition Amplitude Model (AAM). Recently, FreMEn was used to build a map that considered not only the human presence but also their direction [63], [64]. Such a map provides information that allows for planning navigation in a crowd while adhering to human flows [23]. A nontraditional application of FreMEn can be found in [65], where FreMEn ensured the optimal task division between multiple robots operating in a human-populated environment.

FreMEn proved its superiority to the classic robotic approaches in mapping for navigation and provided original insights into the dynamics of the environment. The question of whether it is possible to utilise FreMEn's qualities in exploration tasks arose. The initial experiments on life-long spatio-temporal exploration provided a hypothesis that the best time and place to explore correlates with the uncertainty of FreMEn [66]. That led to developing the information-based Monte-Carlo scheduler for spatio-temporal exploration [10]. However, visiting places when the robot is still determining what to expect can be contrary to the robot's aims, for example, avoiding busy human flows. This contradiction, known as the exploration-exploitation dilemma, was thoroughly studied in [67]. The effort was crowned by defining the life-long spatio-temporal exploration of dynamic environments [12]. Various exploration strategies were later studied over the spatio-temporal-directional maps [64].

The success of FreMEn, when applied to various robotic tasks, led to its integration into the STRANDS system [9]. The STRANDS system controlled two robots deployed to 2 different environments with different roles. One of them was a security robot operating in an office building in the UK. The second one was operating in an Austrian care home with an overall autonomous deployment of 8 months. As the robot had to fulfil its task while simultaneously building and updating its map, the most promising exploration-exploitation strategy was successfully integrated into the STRANDS system. The gradual growth of its performance in real deployments proved the theoretical conclusions [68]. It should be noted that FreMEn, as a mapping, navigation, and planning system, fulfilled the strict requirements of the privacy policy as it did not need visual input.

3 Frequency Map Enhancement

Although many applications use FreMEn, and it is possible to meet with derived words like `fremenisation', we feel that the definition(s) found in the related articles do not describe the method identically to its implementation [69]. As our work proceeds from it and follows the evolution of ideas related to it, we believe it is necessary to define FreMEn accurately and clarify its features.

Generally, FreMEn [5], [66] replace stationary uncertainty models of binary states in robotics maps with functions of time, represented by their frequency spectra. For example, the probability of grid cell occupancy is not modelled by a single probability value, updated only by direct observation. Instead, each occupancy grid cell contains a frequency spectrum-based time series model derived from the NUDFT over previous observations. As the time series decomposition is not used for reconstruction but for prediction, only a few of the most prominent coefficients of the Fourier spectrum are chosen to serve as parameters of the time series model. The number of these prominent coefficients is referred to as the order of model.

The difference between FreMEn's decomposition and the decomposition using NUDFT lies in the ability of FreMEn to incorporate new observations incrementally, thus being able to provide predictions at any time during the robot operation. This feature enables FreMEn's on-the-fly learning from sparse and irregular data [70]. Let us have a set of candidate angular frequencies $\omega_k \in \Omega$ and the model order o. Starting with $n = 0$ observations, FreMEn updates the model parameters by adding a new measurement $s(t)$ in the following way [69], [70]:

- update mean probability

$$\mu \leftarrow \frac{1}{n+1}(n\mu + s(t));$$

- $\forall \omega_k$ update state spectrum

$$\alpha_k \leftarrow \frac{1}{n+1}\left(n\alpha_k + s(t)e^{jt\omega_k}\right);$$

- $\forall \omega_k$ update observation spectrum

$$\beta_k \leftarrow \frac{1}{n+1}\left(n\beta_k + e^{jt\omega_k}\right);$$

- update number of observations

$$n \leftarrow n+1.$$

Then it can perform a prediction at any time t_0 considering all previous measurements:

- $\forall k$ calculate predictive spectrum $\gamma_k \leftarrow \alpha_k - \mu\beta_k$,
- choose o components $\{\gamma_c\}_1^o$ with highest $|\gamma_k|$,
- and predict state $s(t_0)$:

$$s(t_0) = \begin{cases} 1 & \text{if } \mu + 2\sum_1^o |\gamma_c| \cos(t_0\omega_c - \arg(\gamma_c)) > 0.5 \\ 0 & \text{otherwise.} \end{cases}$$

FreMEn was also defined as an estimator of the probability of occupied state occurrence $P(s(t) = 1)$ [5], [10] in a way that the prediction (3) was transformed into:

$$P(s(t_0) = 1) =$$

$$= \min\left(\max\left(\mu + 2\sum_1^o |\gamma_c| \cos(t_0\omega_c - \arg(\gamma_c)), 0\right), 1\right). \tag{3.1}$$

The considered angular frequencies ω_k and the order of the model o have to be chosen wisely in advance. The original advice for the set of candidate angular frequencies was proposed in [10] as $\omega_k \in \Omega_{24} = \{2\pi k/86400\}_{k=1}^{24}$ (note that one day consists of 86400 seconds). The set Ω_{24} served well in robotic applications where the most significant dynamics came from day/night changes, like topological localisation [15] based on the visibility of features in the video frames, or models of human behaviour calculated over few days long datasets that do not cover week routines [10]. However, it is possible to deduce that the set of candidate angular frequencies used by FreMEn includes the angular frequency $2\pi/604800$, for example, from notes in results [66], from the graphs [60], or by finding a back-reference from one article [10] to another [66]. From the associated source code [69], one can uncover that the standard set of candidate angular frequencies in FreMEn and FreMEn-derived methods is $\omega_k \in \Omega_{168} = \{2\pi k/604800\}_{k=1}^{168}$. The set $\Omega_{168} \supset \Omega_{24}$ includes the frequencies once-per-week and once-per-day, which reflects the environmental dynamics derived from the fact that the social system forces the majority of the population to work on a daily and weekly basis [16]. Ω_{168} is the best known and verified in the field set of candidate angular frequencies for robots deployed in human, calendar-based working environments. Deployment of the algorithm in an environment with different nature of underlying dynamics would require manual construction of the frequency set [40] derived from the domain knowledge.

4 Hypertime

The most popular environmental model in robotics is the occupancy grid [71], which is both for localisation and motion planning. Thus, most spatio-temporal models build spatial representations on the occupancy grid paradigm. In [4], the authors model the typical direction of change in every cell based on the previous and current state of the measured phenomenon in the neighbouring cells. Another model can be found in [72], where authors predict the path of the measured phenomenon based on the actual situation in the grid using the input-output Markov model. The long-term model of the changes in the occupancy grid is based on the spectral analysis of occupancy changes in every cell during the long period [5]. The authors then extended this spatio-temporal model to predict also the direction of the movement through the cell in the specific time [63]. The STRANDS project [9] indicated that the periodicities of human habits (sleep, going to work) are dominant compared to trends, and modelling the periodicities is beneficial for robots.

However, the occupancy grid is memory-consuming, resulting in quantisation noise. This drawback led to the first methods representing the space in a continuous domain [27]. Although the continuous models are computationally intensive to build and maintain, specific optimisations can be used to speed up the model building so that they could be applied in robotics [51]. For example, the authors of [4] showed that using continuous models built by expectation-maximisation methods allows them to model the movement of crowds and the flow of the wind [73]. They also showed that the model of the movement of people could be used to improve the efficiency and safety of navigation [74]. As in the case of spatial models, authors like [75] showed that using continuous models of time periodicity results in better performance than dividing the timeline into arbitrary intervals, like hours of a day or days of a week. Inspired by the success of periodic models of time to represent environmental changes and the e ciency of continuous models, we proposed a continuous version of FreMEn, of which the main idea lay in a speci c transformation of the time domain intended to represent the long-term periodical dynamics of human-populated environments [16].

4.1 Concept

In [76], we proposed a concept of modelling human activities over time in their natural environment. We hypothesised that there are some patterns of human behaviour over the timeline. The conventional approach to time series forecasting divides time-dependent events into three components { trend, seasonal and cyclic patterns { and analyses them separately [77]. Cyclic patterns are generally not predictable changes in the time series, seasonal patterns are periodical changes, and the trend is continuous growth or decrease of measured values. However, as these patterns are derived from the routines and habits of humans, we hypothesised that the patterns show periodical nature with no or negligible trends.

We hypothesised that people's di erent habits form time-dependent events that occur on a regular basis with some randomness (morning hygiene, lunch). We need to estimate their distribution parameters to create the model of the periodical time-dependent patterns of human behaviour. However, the timeline unfolds inde nitely, and it is not possible to repeatedly observe random events at the same time. Thus, in nite, continuous linear time is not suitable as a domain for time-dependent feature parameters estimation, especially for short-duration and rare events - the continuous nature of time does not allow for a description of such events in a classical statistical way.

A popular approach in robotics to model periodical events is to create seasonal windows of a prede ned length, usually one day [37], [38]. The distribution of periodical events with a periodicity matching the length of a window is then represented by a histogram over the window. However, seasonal windows produce discontinuity on their borders, which contradicts human behaviour. Let us consider these examples outlined in [18] to explain the periodicity and continuity of human behaviour:

Figure 4.1: An example of the Hypertime projection. Positive detections during three days are summed up and projected into a Hypertime with the one-day periodicity. The parameters of the distribution of a random time-dependent phenomenon that exhibits a periodic behaviour can be easily estimated.

- the human behaviour is very similar during every morning as opposed to the di erence in behaviour during the morning and afternoon of one randomly chosen day,
- human behaviour ve minutes before midnight and ve minutes after midnight is probably very similar, although we compare behaviour on two different days,
- and, in contrast, human behaviour during Sunday afternoon probably di ers from behaviour on Monday afternoon.

To ensure the continuity on the borders of the time intervals induced by the natural periodicities, we project the time into a set of circles, where every circle is derived from a periodicity detected in a measured phenomenon. In this way, similar time-dependent events matching the periodicity are projected into similar positions of the circle, causing the distances of periodically-occurring events to be low even if signi cant time intervals separate these events, see Figure 4.1. The projection naturally clusters similar time-dependent events and preserves continuity while ensuring that the domain of time-dependent phenomena is constrained. The measured phenomenon projected into this vector space can then be analysed using standard statistical and machine-learning tools, and therefore, it is possible to estimate the parameters of the time-dependent periodical patterns' distributions.

4.2 Modelling Binary Data

4.2.1 HyT

A proposed method supposes periodical patterns in the observed phenomenon { functions derived from human habits { over the timeline. Prominent periodicities in the data de ne a projection of time into a constrained vector space. In this

Figure 4.2: HyT iterations: The method input are observed occurrences of a given phenomenon over time (top left). Then, a model of the data is established (top right). Then, a dominant frequency of the model error is found by Eq. 4.1, the data points are projected into a unit circle, and their distribution is modelled (bottom left). The model is then compared to the original data (bottom right), and the process is repeated until the model error decreases.

vector space, we estimate distributions of patterns in human behaviour, providing a model of human customs.

The most influencing periodicities in the data can be detected using the Spectral analysis [78], and parameters of the mixture of distributions can be gathered using the Expectation-Maximisation model for Gaussian mixtures [79]. Thus, the presented method combines spectral analysis and clustering. The method has to determine the number and length of the periodicities from the non-uniformly sampled data. Doing so in one step would be susceptible to spectral aliasing [80] with the consequence of finding wrong periodicities. Thus, we choose an iterative approach that comprises three steps:

1. identification of the most prominent periodicity,

2. projection of time into more dimensional vector space, and consequent clustering of occurrences of the phenomenon over the vector space,

3. and building a model.

Identification of the Most Prominent Periodicity

Let us have time series $R(t_i)$, $i = 1 \dots n$, where $R(t_i) = 1$ for detected and $R(t_i) = 0$ for not detected occurrence of the studied phenomenon in the time t_i. Let the function $M(t_i)$ be a default model, an estimation of the expected value of $R(t_i)$: $M(t_i) = \frac{1}{n}\sum_{i=1}^{n} R(t_i)$.

First, we apply the spectral decomposition derived from non-uniform Fourier transform[5], [81], [82] on the difference between this time series and the model to find prominent periodicities. In particular, for every considered period P_k, we calculate components of the frequency spectrum and select the most prominent periodicity P_P as follows:

$$P_P = \arg\max_{P_k} \frac{1}{n}\sum_{i=1}^{n} (R(t_i) - M(t_i)) e^{(-1j)2\pi t_i / P_k}; \qquad (4.1)$$

where $M(t_i)$ is the actual model of the time series.

Projection and Clustering

Then, for the chosen period P_P we project occurrences of phenomenon into a 2d vector space (Figure 4.2) as follows:

$$t_i \; ! \quad \left(\cos \frac{2 t_i}{P_P} ; \sin \frac{2 t_i}{P_P} \right) ; \tag{4.2}$$

where $\left(\cos \frac{2 t_i}{P_P} ; \sin \frac{2 t_i}{P_P}\right)$ forms a circle in 2d plane which represent the periodicity and continuity of the occurrences. The time values of occurrences with a similar position relative to periodicity P_P are projected on a similar position on the circle; see Figure 4.2. Then, we apply clustering over the projected occurrences. We call the projection of time into the multiple circles Hypertime.

Model Building

We assume that the time-dependent occurrences of the phenomenon projected into the Hypertime, denoted as $@t_i$, are distributed in a way which allows modelling their distribution by Gaussian mixtures. By building the mixture models of occurrences $GMM(@t_i)$, we obtain a model characterised by a membership $u_{i;j}$, cluster centres c_j and covariances $_j$. These allow us to determine the probability that a given projected sample $@t_0$ belongs to a particular cluster using a ^{2} distribution:

$$Pr_{0;j} = 1 \quad Pr \left[Q \quad @t_0 \quad c_j \right)^T \; _j^{\,1} (@t_0 \quad c_j) \right] ; \tag{4.3}$$

where $Q \quad ^2(d \quad 1)$ and d is dimensionality of the constructed vector space. Then the prediction of the expected value of a Bernoulli random variable at a given time, the occurrence at time t_0, is given by the following equation:

$$E[R(t_0)] = \sum_{j=1}^{c} u_{0;j} \, Pr_{0;j} : \tag{4.4}$$

The model $M(t_i)$ for equation (4.1) in the next iteration is calculated as follows:

$$M(t_i) = \sum_{j=1}^{c} u_{i;j} \, Pr_{i;j} \, : \tag{4.5}$$

Repeating the Process

Once we know the new $M(t_i)$, the process is repeated using Eq. 4.1, and each iteration extends the vector space with another couple of dimensions. We denote these additional projections of t_i as ${}^{@}t_i$; where:

$$ {}^{@+1}t_i = \quad {}^{@}t_i ; \cos\frac{2\ t_i}{P_{P_{@+1}}} ; \sin\frac{2\ t_i}{P_{P_{@+1}}} \quad ; \tag{4.6}$$

in particular ${}^{2}t_i = \ {}^{1}t_i ; \cos\frac{2\ t_i}{P_{P_2}} ; \sin\frac{2\ t_i}{P_{P_2}} = \ \cos\frac{2\ t_i}{P_{P_1}} ; \sin\frac{2\ t_i}{P_{P_1}} ; \cos\frac{2\ t_i}{P_{P_2}} ; \sin\frac{2\ t_i}{P_{P_2}} \ ;$ and ${}^{1}t_i = \ \cos\frac{2\ t_i}{P_{P_1}} ; \sin\frac{2\ t_i}{P_{P_1}}$. The iterations stop, and the final model is chosen when the model reaches chosen criterion – more on choosing the proper criterion can be found in the following part, Section III. The whole method is denoted by the abbreviation HyT.

4.2.2 HyTS

The previous method modelled a given phenomenon over time domain only. However, in [16], we also hypothesised the possibility of extending traditional spatial models with the Hypertime projection. We call this projection Hypertime Space, and the method HyTS.

Let us have spatio-temporal detections of occurrences and non-occurrences $f(x_i; t_i)g_{i=1}^{n}$ constituting spatio-temporal series $R(x_i; t_i)$. Let us have a default model $M(x_i; t_i) = \frac{1}{n}\sum_{1}^{n} R(x_i; t_i)$: Similarly to Equation (4.1), we calculate the most prominent periodicity P_P:

$$P_P = \arg\max_{P_k} \frac{1}{n} \sum_{i=1}^{n} (R(x_i; t_i) \quad M(x_i; t_i)) \, e^{(\ 1j)2\ t_i = P_k} : \tag{4.7}$$

Then we project every measurement $(x_i; t_i)$ into the new vector space as follows:

$$(x_i; t_i) \; ! \quad \left(x_i; \cos\frac{2 t_i}{P_P}; \sin\frac{2 t_i}{P_P}\right); \tag{4.8}$$

denoted as ${}^{@}x_i$: Similarly to the previous case, we perform clustering $GMM\,({}^{@}x_i)$ over the extended space to obtain the model and recalculate the model:

$$M\,(x_i; t_i) = \sum_{j=1}^{c} u_{i;j} \, Pr_{i;j}\,; \tag{4.9}$$

where

$$Pr_{i;j} = 1 \quad Pr\left[Q \quad ({}^{@}x_i \quad c_j)^{T} \;_j{}^{1}({}^{@}x_i \quad c_j)\right]: \tag{4.10}$$

This allows us to repeat the process, extend the vector space with two additional dimensions, and create a more dimensional Hypertime Space, where:

$${}^{@+1}x_i = \left({}^{@}x_i; \cos\frac{2 t_i}{P_{P_{@+1}}}; \sin\frac{2 t_i}{P_{P_{@+1}}}\right); \tag{4.11}$$

and ${}^{0}x_i = x_i$; ${}^{1}x_i = \left(x_i; \cos\frac{2 t_i}{P_{P_1}}; \sin\frac{2 t_i}{P_{P_1}}\right)$; etc.

4.2.3 People Density Prediction over Occupancy Grid

One of the most desired spatial models used in robotics is an occupancy grid, and, therefore, our model should be able to predict the occupancy of grid cells at a particular time. For that, one has to determine a basic volume element, which defines the resolution (granularity) of the discrete model.

The volume b of the basic element of the space-time has to be chosen accordingly to the purposes of the model (for example, 1 squared meter hour $[m^2 h]$). Thus, we define a histogram H with bins of volume b. Spatio-temporal positions of bins are defined by vectors $(b_h; t_h)$, where h are meaningful indices. Vectors $(b_h; t_h)$ should lie inside the area of bins, preferably in the spatio-temporal centres of bins. The values v_h connected to each bin are then calculated as a sum of values $R\,(x_i; t_i)$ assigned to every $(x_i; t_i)$ that lies inside of the volume of the corresponding bin. Using the histogram H, we modelled the most likely number of occurrences of the measured phenomenon per chosen volume unit b as follows:

1. Assumming that every measure $R(x_i; t_i)$ is one or zero, apply a clustering method to $^{@}x_i$ for every $(x_i; t_i) : R(x_i; t_i) = 1$,

2. using a given clustering method, calculate centroids c_j and covariance matrices $_j$ of every cluster, and membership $u_{i;j}$ of every $^{@}x_i$ to every cluster, where j is index of clusters,

3. using c_j and $_j$ calculate membership $u_{h;j}$ of every $^{@}b_h$ to every cluster,

4. calculate cluster weights $_j$ as

$$ {}_j = \frac{\sum_i u_{i;j}}{\sum_h u_{h;j}} \tag{4.12} $$

.

5. Then the function

$$ (x_0; t_0) = \sum_{j=1}^{c} {}_j \, u_{0j} \, P r_{0;j} \tag{4.13} $$

estimates the number of occurrences of the observed phenomenon in the neighbourhood with volume b around $(x_0; t_0)$.

4.2.4 Metaparameters

The quality of the model is based on two parameters, the number of clusters and the set of chosen periodicities to create the Hypertime. Typically, we create several models with di erent parameters and choose the one with minimal error. Unfortunately, we did not nd any elegant heuristic to estimate these parameters. However, using only one cluster when building HyT simpli es the algorithm computationally and, in most cases, such models provide comparable (and sometimes better) results.

4.3 Modelling Data that Include Only Occurrences

The proposed method can be applied to the data that does not need information about events that did not happen - non-occurrences, $s(t_i) = 0$. It is beneficial for higher-dimensional tasks like building maps of human flows, where it is pretty memory and computationally demanding to create an exhaustive set of possible values that never happened. Moreover, some sensors, like LiDARs, and some detection methods, like person detection and localisation method [83], do not produce non-occurrences - and the proposed method can be applied directly to the data without its preprocessing.

The training process is similar to those modelling the binary data, except the step *Identification of the Most Prominent Periodicity* in Section 4.2.2. We do not know the whole spatio-temporal series $R(x_i; t_i)$. Therefore, we need to define a histogram H with bins $(b_h; t_h)$ of meaningful volume b and calculate values v_h connected to each bin as a number of occurrences belonging to each bin. Then we define spatio-temporal series $R(b_h; t_h)$:

$$R(b_h; t_h) = \frac{v_h \; t_{min} \; |\{(b_h; t_h)\}|}{t_{whole}}; \tag{4.14}$$

where t_{min} is an estimation of minimal time between consecutive measurements derived from a sensor frequency of gauging, t_{whole} the length of time the sensor collected the data, and $|\{(b_h; t_h)\}|$ is a number of bins in histogram H. The ratio $\frac{t_{whole}}{t_{min}}$ can be understood as a theoretical maximum of occurrences the sensor can detect, and, therefore, equation (4.14) normalises the number of occurrences in each bin to a value between 0 and 1. That allow us to directly apply actual model $M((x); t)$ into a calculation the most prominent periodicity P_P:

$$P_P = \arg\max_{P_k} \frac{1}{n} \sum_{i=1}^{n} (R(b_h; t_h) - M(b_h; t_h))\, e^{(-1j)2 t_h/P_k}. \tag{4.15}$$

4.3.1 Note on Size of Bins

Although the size of volume b impacts the computational complexity, it looks like the impact on the quality of a model is insignificant when the temporal size of the bins is between t_{min} and half of the shortest considered periodicity. Therefore, I would recommend half of the shortest considered periodicity as a default setting – usually half an hour. I usually set the spatial size up to one meter when modelling human-centred data, as this looks meaningful considering human personal space. However, the size of the bins was not tested exhaustively, and researchers should take this statement with caution.

4.4 Regression-Like Models

In previous sections, 4.2 and 4.3, we expected that the measured value would reflect not only a position in space but also a fact that the phenomenon occurred. There are phenomena like temperature, pressure, or the number of people in an area that can be measured anytime and always provide some actual value. We cannot understand a measurement as an occurrence of the phenomenon - it is an occurrence of gauging. In such a case, we know that the occurrence of the phenomenon is a certain event, while gauging occurs randomly in time.

The training process is similar to the one described in a section about modelling data that includes only occurrences of a phenomenon, Section 4.3. However, as we expect heteroscedasticity and multimodality of the variable v, the final model $M^F(t)$ needs to provide a robot with an estimation of the conditional distribution of variable v in any specific time t_0 (Figure 4.3):

$$M^F(t_0) = M_{v|t=t_0}(v;t) : \tag{4.16}$$

Figure 4.3: Left: Regression method overview. The data points (a,t) observed over time (top, black) are first processed by frequency analysis to determine a dominant periodicity T. Then, the time t is projected onto a Hypertime Space 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, centre, blue) to estimate the distribution of a over the Hypertime Space (bottom, right, green). Projection of the distribution back to the linear time domain allows us to calculate the conditional distribution estimation of a for any past or future time (top, green). Courtesy of [5]. Right: Conditional distribution detail. Courtesy of [84].

4.5 Use Cases

4.5.1 Anomaly Detection

This use case was motivated by the security scenario of the STRANDS project [9], where a regularly-patrolling robot observes the presence of people in a given area and reports anomalous behaviour. A robot needs to consider the temporal context of a given activity or person's presence. For example, being present during late hours in an office is an unusual event. In contrast, people's occurrence in an office during the day is common.

Figure 4.4: Longer training dataset. Three weeks of detections of one person in his office. Time series values acquire 1 when the person is detected and 0 otherwise.

Dataset

We evaluated HyT on the real-world dataset from Lincoln University. It consists of twenty weeks of detection of one person in his office. We divided this dataset into the training part, which consists of the first three weeks (Figure 4.4), and the test part, which consists of the eighteenth week (Figure 4.5). The eighteenth week was chosen intentionally - it contains a relatively high number of anomalous events. We labelled the outliers in the test dataset manually. Namely three situations: on Tuesday evening, the human subject returned after leaving his office; on Wednesday evening, the subject worked until night; and the last labelled situation was on Saturday when he was working on a non-working day. From the training dataset, we created twenty-one new training datasets. The first consists

Figure 4.5: Test dataset. The eighteenth week of detections with highlighted outliers (black) on Tuesday, Wednesday, and Saturday. The outliers were labelled manually.

of the first day's measurements, the second one consists of the first two days' measurements, and so on. Using such a set of training datasets, we can compare the gradual improvement of the outlier detection ability and speed of the learning of the studied methods.

Methods

We compare our method, HyT, with four different methods. The first of them is Prophet, an open source time series forecasting tool created by Facebook [85], the second is FreMEn [5] with a model order $o = 5$, and the last two are histograms Hist24 and Hist168. Hist24 calculates the mean of occurrences in every hour of all days in a training dataset, and Hist168 calculates one-hour means over the whole

Figure 4.6: Evolution of Matthews correlation coefficient using $L = 0.90$. On the x-axis are the numbers of days used to train the models; on the y-axis are the values of Matthews correlation coefficient $\langle -1, 1\rangle$. A coefficient of value 1 means correct labelling of outliers by the corresponding method.

week. Note that the chosen time windows correspond to the most prominent periodicities found by FreMEn applied to the whole (training and test together) dataset. Contrary to histograms, HyT and FreMEn automatically calculate the most prominent periodicities from individual training datasets.

Methodology

To quantify the outlier detection ability of the tested methods, we chose the Matthews correlation coefficient [86], which is suitable to measure the quality of binary classifications, in our case, inliers and outliers. We created models for all of 21 training datasets consisting of data from the first day, the first two days, ..., the first twenty-one days of measurements (Figure 4.4), and detected outliers in the 18th week of measurements (Figure 4.5) using the models' prediction of the

Figure 4.7: Evolution of Matthews correlation coefficient using $L = 0.99$. On the x-axis are the numbers of days used to train the models, and on the y-axis are the values of Matthews correlation coefficient $\langle -1; 1\rangle$. A coefficient of value 1 means correct labelling of outliers by the corresponding method.

behaviour of the studied subject. The values of Matthews correlation coefficients for different methods and different training datasets ordered by the length of these datasets for $L = 0.90$ and $L = 0.99$ are shown in Figure 4.6 and Figure 4.7 respectively.

Results

We can see in Figure 4.6 and Figure 4.7 that HyT needed a shorter time to learn features of the studied subject behaviour than the original FreMEn. The ability of HyT to detect anomalous events in the test dataset is significantly better than FreMEn's. Compared to the other methods, HyT exhibits robustness to the choice of significance level. Moreover, HyT is better than the popular robotics technique, the histograms, even if these histograms use predefined periodicities. Hist168 with

$L = 0{:}99$ is able to maximise its ability to predict outliers. However, it needed two and a half weeks to train fully. HyT was adequately trained during the rst ten days. With its default setting, Prophet represents up-to-date commercial tools for time series analysis. It was unable to learn the model as fast as the others. It should be noted that Prophet could predict outliers similarly to FreMEn when the models were trained over ve and more weeks. Contrary to other tested methods, Prophet models a trend. This ability, however, did not give it an edge in this kind of scenario. Based on our experiments, the assumption that the trend can be neglected for the human behaviour analysis over several weeks was not disproved.

4.5.2 Pedestrians Density

Dataset

To evaluate our method, we utilised a dataset of people's presence collected at the University of Lincoln. The data were collected by a mobile robot equipped with a Velodyne 3D laser range nder. The robot was driven to a location which provided a good overview of one of the T-shaped corridor junctions (Figure 4.8). To localise people in the measurements provided by the laser 3D scanner, we used a reliable and e cient person detection and localisation method [83]. Since the robot batteries need to be recharged on a daily basis, we could not collect the data in a continuous, 24/7 manner. We had to remove the robot from the observation spot every night when the building was vacant, and no people were in the corridors. The collected dataset contains detections from early mornings to late evenings on weekdays over several weeks. A typical day contains approximately 32000 people detection measurements, which correspond to a large number of people walking or standing in the monitored corridors. The method [83] provides human detection results as a single vector, e.g., a point in space. However, for mobile robot path planning, a human represents an obstacle with a particular spatial volume. For this reason, we preprocessed the dataset by substituting every detection vector with a set of vectors in its spatial neighbourhood with a diameter of 0.5m.