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. Tomáš Krajník, Ph.D.
Department of Computer Science
Faculty of Electrical Engineering
Czech Technical University in Prague
Technická 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
Technická 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. Tomáš 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 robotů se očekává, že budou pomáhat lidem s mnoha různými úkoly. Předpokládáme, že kromě schopnosti řešit úkol se servisní roboti potřebují také začlenit do společnosti, aby poskytovali službu přijatelným způsobem. Lidská společnost má pravidla, která se řídí pracovními rutinami, tradicemi a odvozenými zvyky. Porušení pravidel jakoukoliv entitou, včetně autonomních robotů, kazí psychosociální atmosféru a vede k následnému vyloučení nepřizpůsobivé entity – v případě robota k odmítnutí jeho pomoci.

V této práci navrhujeme způsob, jak by se autonomní obslužný robot mohl alespoň začlenit do lidské společnosti – aplikováním časoprostorových map umožňujících předpovědi dynamiky vyvolané člověkem do jeho plánovacího modulu. Definovali a vyhodnocovali jsme časoprostorové modely vhodné pro autonomní servisní roboty v různých scénářích. Navrhli jsme metodiku pro definování různorodého souboru kritérií pro hodnocení a porovnávání různých časoprostorových map z hlediska použitelnosti v konkrétních robotických úlohách. Odvodili jsme specifické kritérium pro vyhodnocení časoprostorových map pro rozvržení a plánování trajektorií s ohledem na lidi. Pomocí tohoto kritéria jsme porovnali vyčerpávající sbírku zjednodušených přístupů k časoprostorovému mapování, které čtenáři poskytly náhled na jejich odlišnosti, a současných nejmodernějších metod, včetně té navrhované, poskytující komplexní posouzení jejich silných a slabých stránek. Hodnotili jsme také dopad integrace navržené časoprostorové mapy do plánovacího modulu v polním robotickém experimentu, který analyzoval reakce neinformovaných lidských subjektů.

Výsledky polního experimentu ukázaly radikální rozdíl v reakcích na robota, který byl a nebyl schopen předpovídat lidské toky. Robot, který nepoužíval časoprostorovou mapu, některé lidi naštval a jednoho člověka rozzlobil, zatímco robot schopný následovat lidské toky svůj úkol nepozorovaně vyřešil. Navržená metoda časoprostorového modelování úspěšně konkurovala svým konkurentům. Své kvality prokázala při integraci do veřejné aplikace. Poskytla předpovědi pokrývající celou zemi z velmi řídkých údajů.

Klíčová Slova: časoprostorové prognózování, dlouhodobá autonomie, plánování úkolů s ohledem na člověka, porovnávání časoprostorových 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 artificial intelligence, society expects robots to help people with many different 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 flow 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 incorpo-

rate 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 benefit 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 efficiency 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 efficiency 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 efficiency [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-specific maps [26], [37]–[41]. Another approach is based on kernel warping. The map includes a continuous Gaussian mixture model with predefined periodical kernels [42]–[46]. There also exist approaches of manual preprocessing the data to find 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 flow. 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 define 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

$$\mu \leftarrow \frac{1}{n+1}\left(n\mu + s(t)\right),$$

- $\forall \omega_k$ 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 $s(t_0)$:

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

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

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

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 efficiency of continuous models, we proposed a continuous version of FreMEn, of which the main idea lay in a specific 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 different 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 indefinitely, and it is not possible to repeatedly observe random events at the same time. Thus, infinite, 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 predefined 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:

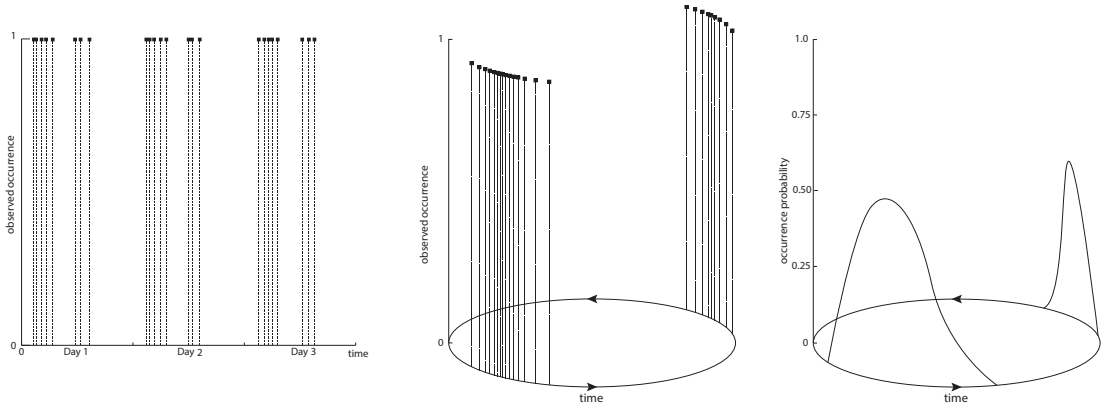

Figure 4.1: An example of the Hypertime projection. Positive detections during three days are summed up and projected into a Hypertime with the one-day periodicity. The parameters of the distribution of a random time-dependent phenomenon that exhibits a periodic behaviour can be easily estimated.

- the human behaviour is very similar during every morning as opposed to the difference in behaviour during the morning and afternoon of one randomly chosen day,
- human behaviour five minutes before midnight and five minutes after midnight is probably very similar, although we compare behaviour on two different days,
- and, in contrast, human behaviour during Sunday afternoon probably differs 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 significant 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 define a projection of time into a constrained vector space. In this

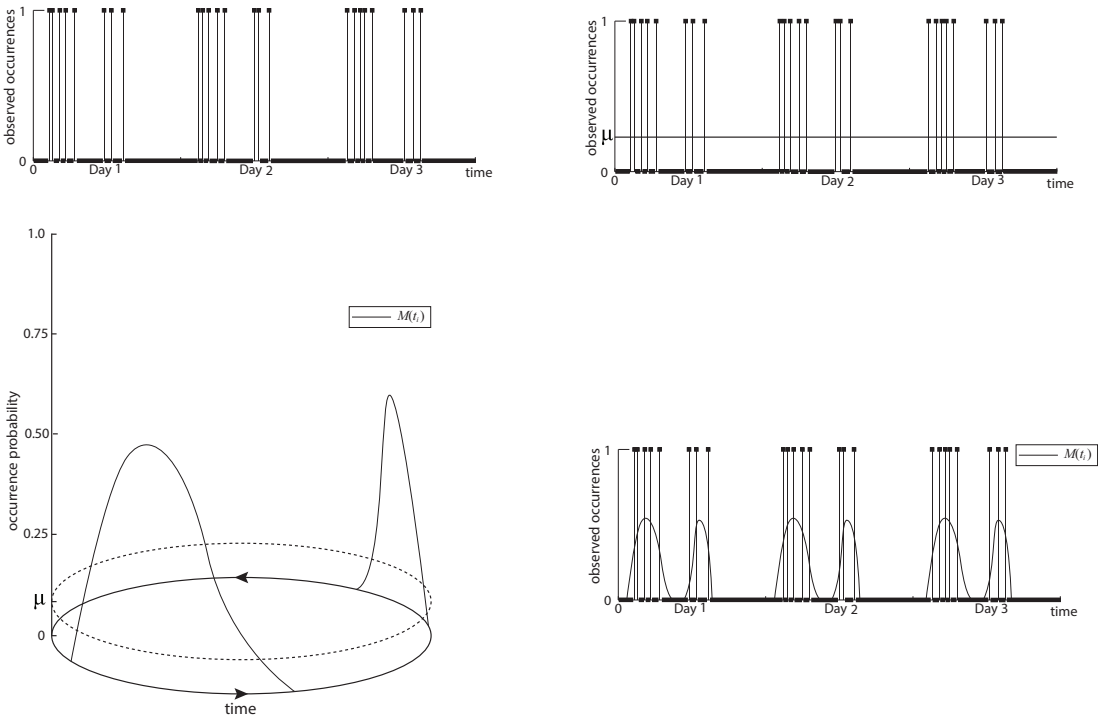

Figure 4.2: HyT iterations: The method input are observed occurrences of a given phenomenon over time (top left). Then, a model of the data is established (top right). Then, a dominant frequency of the model error is found by Eq. 4.1, the data points are projected into a unit circle, and their distribution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

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

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:

$$t_i \rightarrow \left(\cos\frac{2\pi t_i}{P_P}, \sin\frac{2\pi t_i}{P_P}\right), \tag{4.2}$$

where $\left(\cos\frac{2\pi t_i}{P_P}, \sin\frac{2\pi t_i}{P_P}\right)$ 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 ${}^{@}\mathbf{t}_i$, are distributed in a way which allows modelling their distribution by Gaussian mixtures. By building the mixture models of occurrences $GMM({}^{@}\mathbf{t}_i)$, we obtain a model characterised by a membership $u_{i,j}$, cluster centres $\mathbf{c}_j$ and covariances Σ_j. These allow us to determine the probability that a given projected sample ${}^{@}\mathbf{t}_0$ belongs to a particular cluster using a χ^2 distribution:

$$Pr_{0,j} = 1 - Pr\left[Q\left({}^{@}\mathbf{t}_0 - \mathbf{c}_j\right)^T \Sigma_j^{-1}({}^{@}\mathbf{t}_0 - \mathbf{c}_j)\right], \tag{4.3}$$

where $Q \sim \chi^2(d-1)$ 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:

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

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 ${}^{@}\mathbf{t}_i$, where:

$${}^{@+1}\mathbf{t}_i = \left({}^{@}\mathbf{t}_i, \cos\frac{2\pi t_i}{P_{P_{@+1}}}, \sin\frac{2\pi t_i}{P_{P_{@+1}}} \right), \tag{4.6}$$

in particular ${}^{2}\mathbf{t}_i = \left({}^{1}\mathbf{t}_i, \cos\frac{2\pi t_i}{P_{P_2}}, \sin\frac{2\pi t_i}{P_{P_2}} \right) = \left(\cos\frac{2\pi t_i}{P_{P_1}}, \sin\frac{2\pi t_i}{P_{P_1}}, \cos\frac{2\pi t_i}{P_{P_2}}, \sin\frac{2\pi t_i}{P_{P_2}} \right)$, and ${}^{1}\mathbf{t}_i = \left(\cos\frac{2\pi t_i}{P_{P_1}}, \sin\frac{2\pi t_i}{P_{P_1}} \right)$. 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 $\{(\mathbf{x}_i, t_i)\}_{i=1}^{n}$ constituting spatio-temporal series $R(\mathbf{x}_i, t_i)$. Let us have a default model $M(\mathbf{x}_i, t_i) = \frac{1}{n}\sum_{1}^{n} R(\mathbf{x}_i, t_i)$. Similarly to Equation (4.1), we calculate the most prominent periodicity P_P:

$$P_P = \arg\max_{P_k} \frac{1}{n} \sum_{i=1}^{n} \left(R(\mathbf{x}_i, t_i) - M(\mathbf{x}_i, t_i) \right) e^{(-1j)2\pi\, t_i/P_k}. \tag{4.7}$$

Then we project every measurement $(\mathbf{x}_i, t_i)$ into the new vector space as follows:

$$(\mathbf{x}_i, t_i) \rightarrow \left(\mathbf{x}_i, \cos\frac{2\pi t_i}{P_P}, \sin\frac{2\pi t_i}{P_P}\right), \tag{4.8}$$

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

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

where

$$Pr_{i,j} = 1 - Pr\left[Q\left({}^{@}\mathbf{x}_i - \mathbf{c}_j\right)^T \Sigma_j^{-1}({}^{@}\mathbf{x}_i - \mathbf{c}_j)\right]. \tag{4.10}$$

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

$${}^{@+1}\mathbf{x}_i = \left({}^{@}\mathbf{x}_i, \cos\frac{2\pi t_i}{P_{P_{@+1}}}, \sin\frac{2\pi t_i}{P_{P_{@+1}}}\right), \tag{4.11}$$

and ${}^{0}\mathbf{x}_i = \mathbf{x}_i$, ${}^{1}\mathbf{x}_i = \left(\mathbf{x}_i, \cos\frac{2\pi t_i}{P_{P_1}}, \sin\frac{2\pi t_i}{P_{P_1}}\right)$, 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 $[m^2h]$). Thus, we define a histogram $\mathbf{H}$ with bins of volume b. Spatio-temporal positions of bins are defined by vectors $(\mathbf{b}_h, t_h)$, where h are meaningful indices. Vectors $(\mathbf{b}_h, t_h)$ 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 $R(\mathbf{x}_i, t_i)$ assigned to every $(\mathbf{x}_i, t_i)$ that lies inside of the volume of the corresponding bin. Using the histogram $\mathbf{H}$, we modelled the most likely number of occurrences of the measured phenomenon per chosen volume unit b as follows:

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

2. using a given clustering method, calculate centroids $\mathbf{c}_j$ and covariance matrices Σ_j of every cluster, and membership $u_{i,j}$ of every ${}^{@}\mathbf{x}_i$ to every cluster, where j is index of clusters,

3. using $\mathbf{c}_j$ and Σ_j calculate membership $u_{h,j}$ of every ${}^{@}\mathbf{b}_h$ to every cluster,

4. calculate cluster weights α_j as

$$\alpha_j = \frac{\sum_i u_{i,j}}{\sum_h u_{h,j}} \tag{4.12}$$

   .

5. Then the function

$$\rho(\mathbf{x}_0, t_0) = \sum_{j=1}^{c} \alpha_j \, u_{0j} \, Pr_{0,j} \tag{4.13}$$

   estimates the number of occurrences of the observed phenomenon in the neighbourhood with volume b around $(\mathbf{x}_0, t_0)$.

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 different parameters and choose the one with minimal error. Unfortunately, we did not find any elegant heuristic to estimate these parameters. However, using only one cluster when building HyT simplifies 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 $R(\mathbf{x}_i, t_i)$. Therefore, we need to define a histogram $\mathbf{H}$ with bins $(\mathbf{b}_h, t_h)$ 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 $R(\mathbf{b}_h, t_h)$:

$$R(\mathbf{b}_h, t_h) = \frac{v_h \, \Delta t_{min} \, |\{(\mathbf{b}_h, t_h)\}|}{\Delta t_{whole}}, \tag{4.14}$$

where Δt_{min} is an estimation of minimal time between consecutive measurements derived from a sensor frequency of gauging, Δt_{whole} the length of time the sensor collected the data, and $|\{(\mathbf{b}_h, t_h)\}|$ is a number of bins in histogram $\mathbf{H}$. The ratio $\frac{\Delta t_{whole}}{\Delta t_{min}}$ 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:

$$P_P = \arg\max_{P_k} \frac{1}{n} \sum_{i=1}^{n} \left(R(\mathbf{b}_h, t_h) - M(\mathbf{b}_h, t_h) \right) e^{(-1j)2\pi \, t_h / P_k}. \tag{4.15}$$

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 Δt_{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):

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

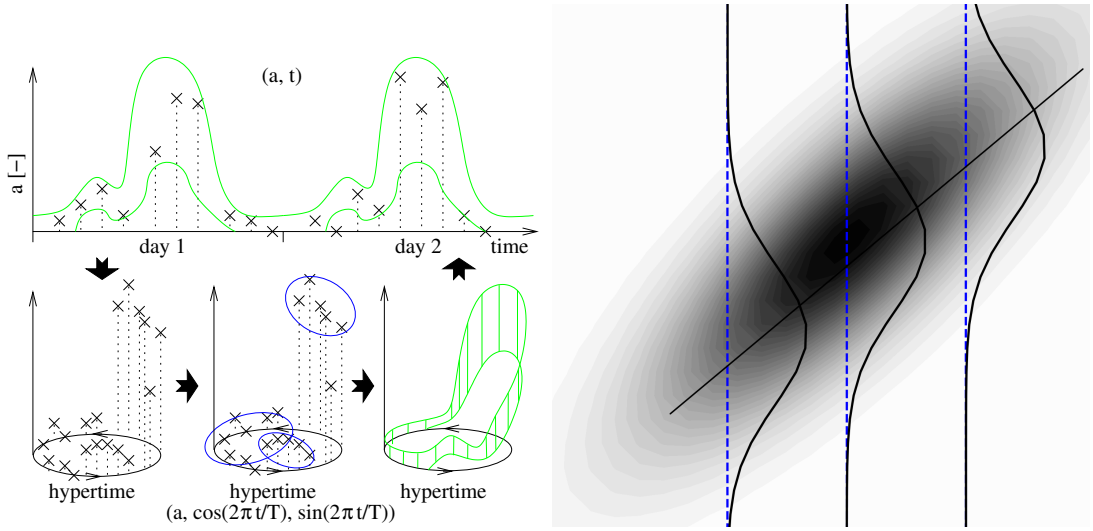

Figure 4.3: **Left:** Regression method overview. The data points (a,t) observed over time (top, black) are first processed by frequency analysis to determine a dominant periodicity T. Then, the time t is projected onto a Hypertime Space and the vectors (a, t) become $(a, cos(2\pi\, t/T), sin(2\pi\, t/T))$ (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.

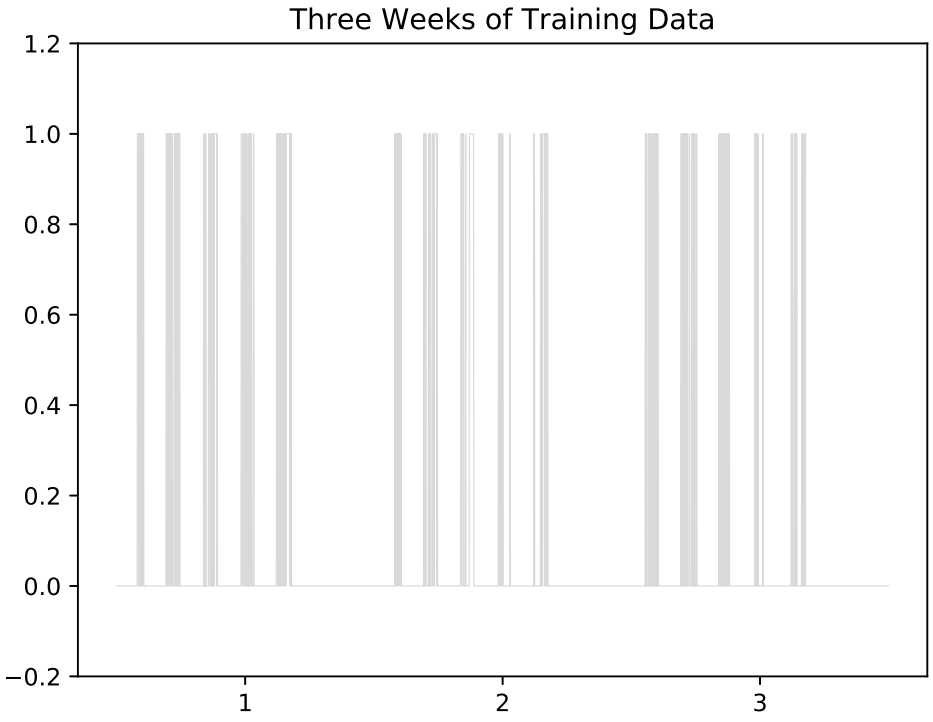

Figure 4.4: Longer training dataset. Three weeks of detections of one person in his office. Time series values acquire 1 when the person is detected and 0 otherwise.

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

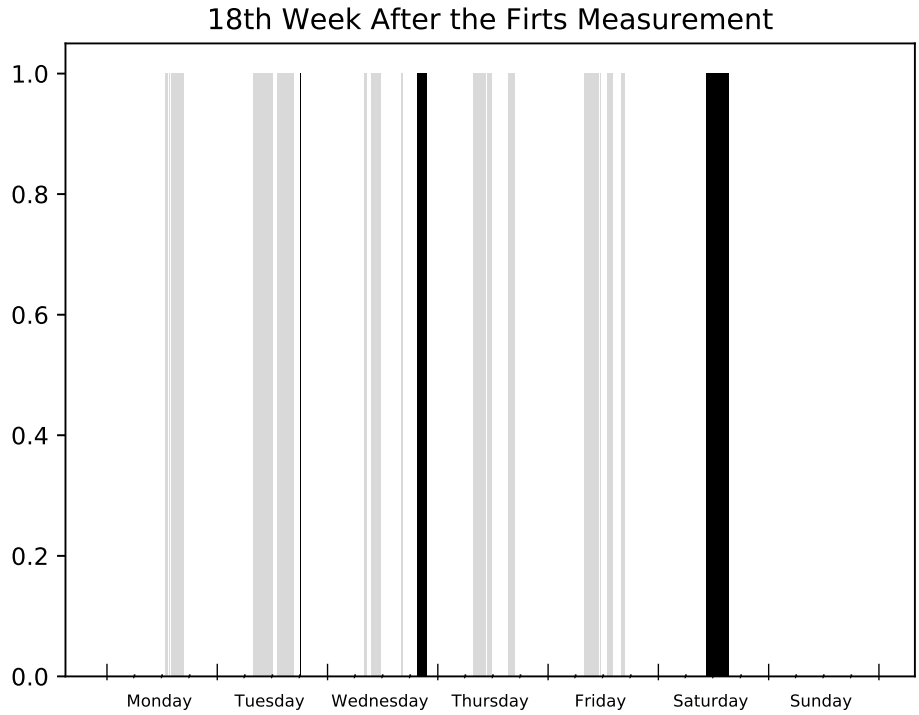

Figure 4.5: Test dataset. The eighteenth week of detections with highlighted outliers (black) on Tuesday, Wednesday, and Saturday. The outliers were labelled manually.

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

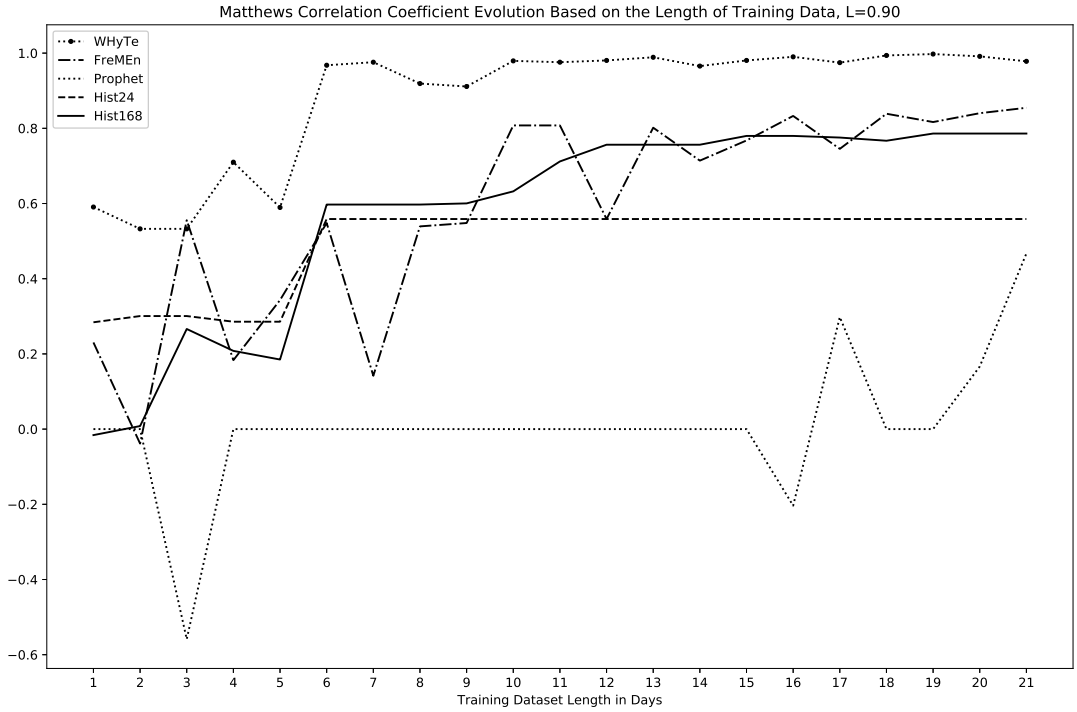

Figure 4.6: Evolution of Matthews correlation coefficient using $L = 0.90$. On the x-axis are the numbers of days used to train the models; on the y-axis are the values of Matthews correlation coefficient $\langle -1; 1 \rangle$. A coefficient of value 1 means correct labelling of outliers by the corresponding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

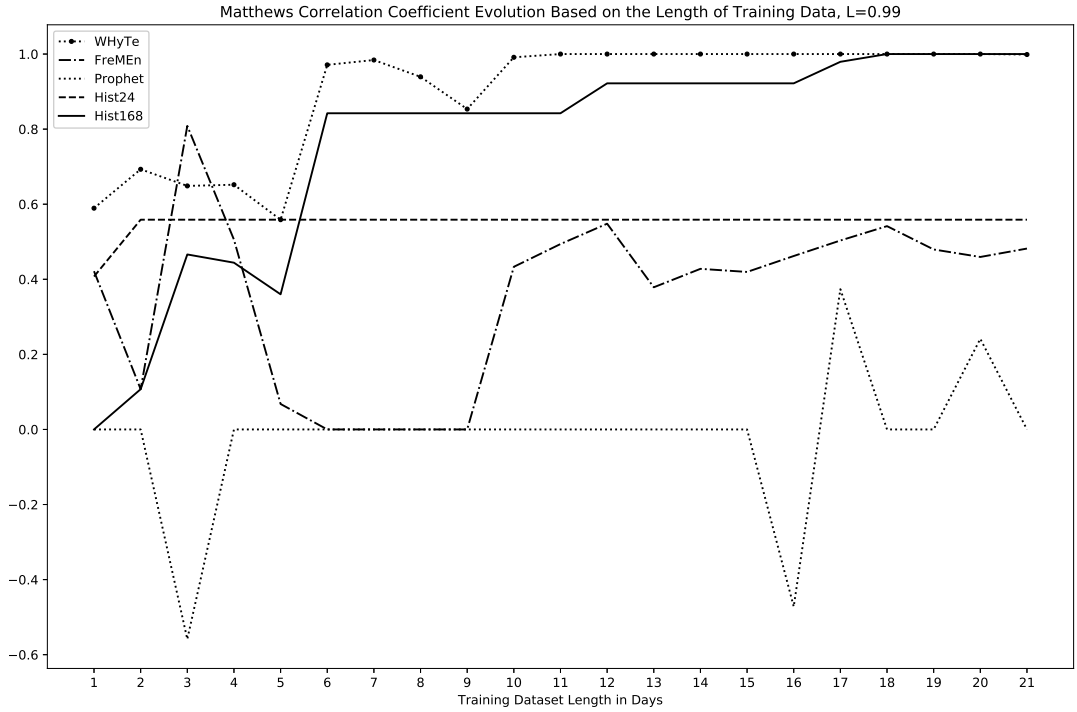

Figure 4.7: Evolution of Matthews correlation coefficient using $L = 0.99$. On the x-axis are the numbers of days used to train the models, and on the y-axis are the values of Matthews correlation coefficient $\langle -1; 1 \rangle$. A coefficient of value 1 means correct labelling of outliers by the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

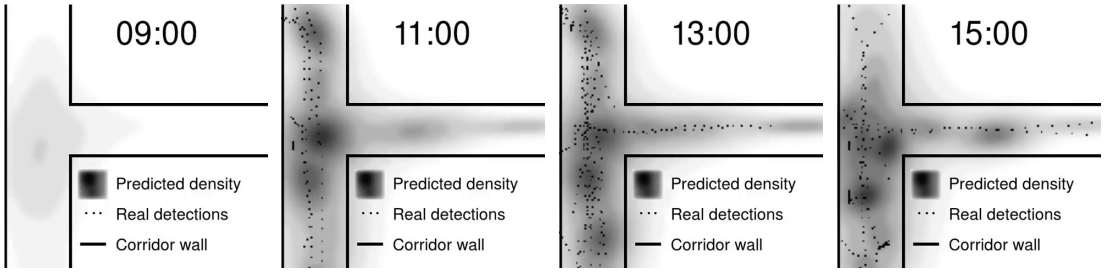

Figure 4.8: Model compared to the data. Black dots represent detections from the test dataset, and grey areas represent the model (darker area means more predicted occurrences). Black lines represent walls of corridors.

Methods

We compare HyTS to three other spatio-temporal models. These models represent the space in a discrete way. In particular, they associate each cell of the spatial 2d grid with a temporal model: Mean, which is simply an average of all past measurements occurring within a given cell, *Hist*, which splits each day into h intervals and predicts the number of occurrences as an average for the relevant time of a day, and FreMEn, which extracts o spectral components from the people occurrence history and uses these periodic components for future predictions. To predict the number of occurrences at a particular location at a particular time, we select the grid cell that corresponds to the location and use the cell's temporal model to perform the prediction. More precisely, to generate the 3-d histogram $\mathbf{M}$ representing a model's spatio-temporal prediction, we first calculate a separate temporal model for each cell of the predefined spatial grid $\mathbf{G}$. Then we create 1-d histogram of each temporal model in every cell and generate a spatio-temporal histogram of predictions $\mathbf{M}$. On the contrary, HyTS represents the space in a continuous, and to generate the histogram $\mathbf{M}$, it simply calculates the cell value according to (4.13). As the corridor is T-shaped, we decided to follow the natural spatial structure of an environment and set $c = 3$ clusters for HyTS.

Methodology

The primary purpose of the evaluation is to estimate the predictive capability of HyTS models. We split the gathered data into training and test sets and only

learned the models from the training set. The training dataset consists of two weeks of measurements, and the test dataset consists of two days of measurements from another week (Wednesday and Thursday). Out of the test data, we created a 3-d, spatio-temporal histograms $\mathbf{H}$ with various basic volumes b ranging from 0.1 to 1.0 meters and 10 to 60 minutes. In particular, we count the number of detections for each bin from the test data.

The error of a method at a specific resolution is calculated as a root-mean-square deviation [87] between histogram $\mathbf{H}$ and matching histogram $\mathbf{M}$:

$$RMSD_{current} = \sqrt{\sum_{bin} (\mathbf{M}_{bin} - \mathbf{H}_{bin})^2}, \tag{4.17}$$

with *bin* defined according to procedure in section 4.2.2. The prediction error differs between different resolutions, Figure 4.9. It was discussed in later work in detail [26].

To better visualise the comparison of prediction capabilities of different methods on different resolutions, we calculate the ratio between the RMSD of prediction of every compared method to the RMSD of prediction gathered from Mean, which can be understood as a representant of the classical occupancy grid:

$$ratio = 1 - RMSD_{method}/RMSD_{Mean}. \tag{4.18}$$

The experimental evaluation is performed by an automated system similar to the one presented in [88]. This system first optimises each method's (HyTS, Hist, and FreMEn) parameters (Hypertime dimension @, number of bins per day h, and model order o). Then it runs pairwise t-tests to determine which of the compared methods performs statistically significantly better than the others.

Table 4.1: Model Size of Compared Methods

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

Results

We can see in Fig. 4.9 that HyTS significantly outperformed the other ones. It is usually more than 5% better than FreMEn and Hist. Fig. 4.8 also demonstrates that the distributions of the people predicted by HyTS depend on time and follow the shape of the corridor even when the algorithm did not force it.

Moreover, the proposed model is much more memory efficient than FreMEn or the other discrete models. This is because our model is continuous, and its complexity is unaffected by the number of cells in the underlying grid. HyTS consists of the positions of clusters $\{\mathbf{c}_j\}$, their weights $\{\alpha_j\}$, covariance matrices $\{\Sigma_j\}$, and a set of used periodicities $\{P_p\}$. With @ = 2 prominent periodicities and $c = 3$ clusters, HyTS occupies 1.7 KiB of memory space - independently of the grid resolution. In contrast, FreMEn needs several numbers for each modelled periodicity in every cell of the grid-based representation. Thus, on our dataset, where the spatial grids have resolutions of 100, 50, 20, and 10 cm, the FreMEn model occupies 1.1, 4.4, 27.2, and 108.9 MiB of memory respectively, see Table 4.1. In addition, the resolution of HyTS is derived from the cluster weights α_j. Therefore it is possible to store different resolutions of one model at almost no memory cost.

The downside of HyTS is its computational complexity, caused by the method being iterative and the growth of dimensionality of Hypertime in every iteration. Comparing runtimes of our implementations, when FreMEn calculates a temporal model of all the grid cells in about a minute, HyTS builds the spatio-temporal model in an hour. Moreover, during the parameters optimisation phase, HyTS has to discretise the time-space to calculate the histogram $\mathbf{M}$, which occupies memory space comparable to the discrete methods (Mean, Hist, and FreMEn). However, once the model building is finished, the histogram is not stored, and the model consists only of the cluster parameters and periodicities.

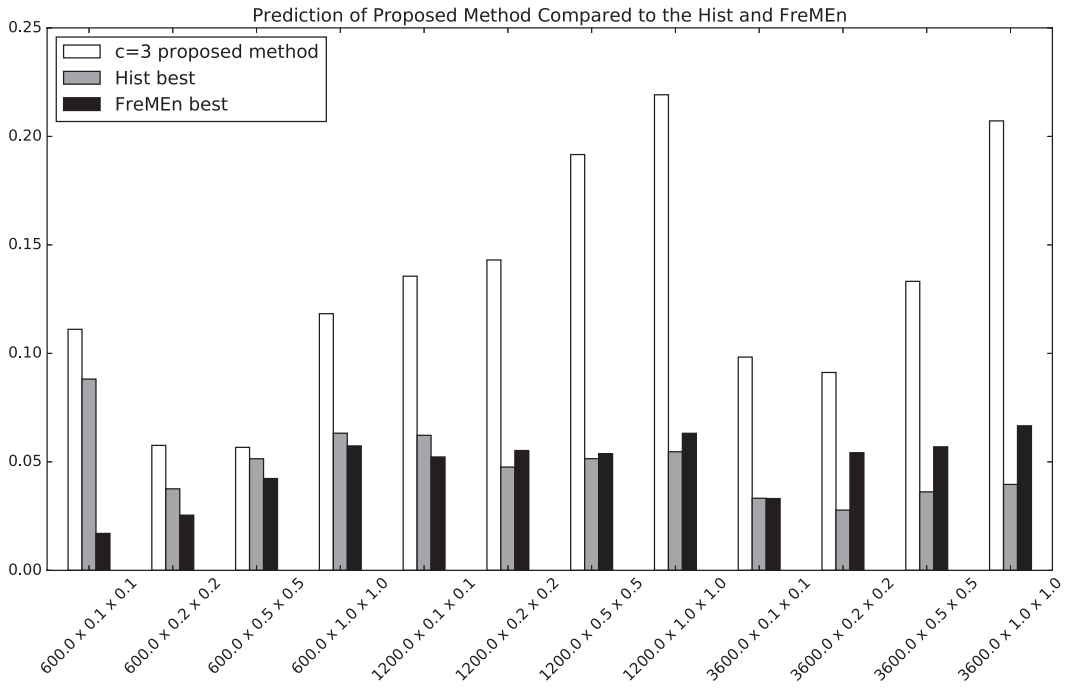

Figure 4.9: Comparison of the predictive power of different methods. The graph shows the prediction error reduction compared to the Mean, which neglects the temporal properties of the people's presence. The indicated values (y-axes) are calculated using Equation (4.18). On the x-axis, there are spatio-temporal sizes of basic volumes. The white bar shows the prediction power of HyTS using three clusters, grey bar Hist and black bar FreMEn.

4.5.3 Pedestrian Flows

Dataset

The approach described above was evaluated using a dataset collected at the Department of computer science at the University of Lincoln. The data recording was performed by a Pioneer 3-AT mobile robot equipped with a 3d lidar (Velodyne VLP-16) and a 2d lidar (Hokuyo UTM-30LX), using a reliable person detection method [83].

During the data collection, the robot remained stationary in a T-shaped junction, which allowed its sensors to scan the three connecting corridors simultaneously, covering a total area of around 75 m^2 (Fig. 4.10). However, since the robot could not stay in the corridor overnight due to safety rules, and it was needed by other researchers occasionally, we did not collect the data on a whole, 24/7 basis. Instead, the data was collected during ten-hour sessions covering the usual working hours. Recharging the batteries was performed overnight when the building was vacant, and no people were on the corridors.

The resulting dataset is composed of 9 data-gathering sessions recorded over four weeks. A typical session contains approximately 30000 detections of people walking in the monitored corridors. Every detection is represented by a vector (t, x, y, ϕ, v) – the detected human's position, orientation, and speed in time. Similar to [27], we added 70000 "no detection" vectors of the positions, orientations, and speeds, where no human was detected (such as random vectors during the night and people walking in the opposite direction than detected ones). As some of the methods in comparison do not model the speed, this value was set to $v = 1.0$ for every measurement. For detailed information about particular methods used in the comparison, see section 4.5.3. The structured overview of the properties of individual methods can be seen in Table 4.2.

The 3d lidar has 16 scan channels with a 360° horizontal and 30° vertical field-of-view and was mounted at the height of 0.8 m from the floor on the top of the robot (Fig. 4.10 left), which allows us to have a perspective that covers the entire environment for data collection (Fig. 4.10 right). All people appearing in the corridor are detected and tracked in the 3d lidar's frame of reference.

More specifically, the 3d point cloud generated by the Velodyne lidar was first segmented into different clusters using an adaptive clustering method [83], then an offline trained SVM-based classifier was used for human classification. The 2d positions of the people are subsequently fed into a robust multi-target tracking system [89] using Unscented Kalman Filter (UKF) and Nearest Neighbour Joint Probability Data Association method (NNJPDA), and the human-like trajectories (in XY-plane) are eventually generated and recorded.

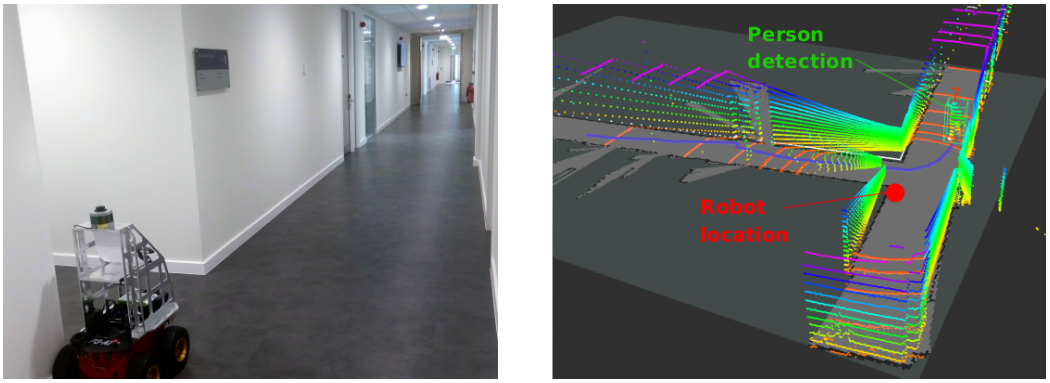

Figure 4.10: Photo of the UoL dataset data collection setup: Robot location in the corridor and example of a person walking as seen by the 3D lidar.

Methodology

Following [90], we divided the dataset into a training and testing subset, where the training dataset consisted of seven days from three weeks, and the test dataset consisted of two days measured out of the time interval of the training dataset. We chose two different criteria to measure the quality of the model. The first is root-mean-square error (RMSE) [87] between model predictions $M(\mathbf{x}_i, t_i)$ and test dataset values $D(\mathbf{x}_i, t_i)$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(M(\mathbf{x}_i, t_i) - D(\mathbf{x}_i, t_i)\right)^2}, \tag{4.19}$$

which is widely used in time series forecasting [91].

The second criterion used is the level of similarity between human motion distributions, occurring at certain times and positions, obtained from the 2 test

days and the ones predicted by the model. This metric is focused on how well the model can predict how a person would move if detected rather than how likely the robot would find one.

In order to do that, we have defined a spatial and temporal grid to cover the entire map and the whole 2 test days. The different approaches can provide a probability motion distribution for any point in time-space. However, obtaining this same distribution with the ground truth data to make the comparison at a single time instance is not possible. The reason is that we do not have enough detections for an instance of time to build a meaningful distribution. Instead, the idea is to compare the distribution obtained from the test data during a defined time interval. In our evaluations, we have used a spatial grid, taking points every 1 meter in x and y directions and 10-minute long time intervals. When there are no detections at nighttime, we assume equal probability for each orientation.

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

$$distance_{map} = \sum_{i=1}^{n} \sum_{b=1}^{k} \frac{(x_b - y_b)^2}{(x_b + y_b)}, \tag{4.20}$$

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

Methods

HyTS Two parameters affect the quality of HyTS - the number of clusters c and the set of periodicities forming the Hypertime. The recent experiments showed that the number of clusters could be relatively small (usually up to 9) [16], and it seems that the number of clusters is related to the topological structure of

the space [20]. For this dataset from T-junction, we chose $c = 3$ clusters. The second parameter can be derived from data iteratively, but recent experiments showed [16], [20], that the quality of prediction does not usually grow with more than @ = 3 added Hypertime 'circles'. We selected the basic set of periodicities as proposed in [5] and discovered three prominent components in the training data (six, twelve, and twenty-four hours), which we used in our method.

STeF-Map STeF-Map [64], which stands for Spatio-Temporal Flow Map, is a representation that models the likelihood of motion directions on a grid-based map by a set of harmonic functions, which capture long-term changes of crowd movements over time. The underlying geometric space is represented by a grid, where each cell contains k temporal models corresponding to k discretised orientations of people's motion through the given cell over time. Since the total number of temporal models, which are of a fixed size, is $k \times n$ where n is the total number of cells, the spatiotemporal model does not grow over time regardless of the duration of data collection. The temporal models, which can capture patterns of people's movement, are based on the FreMEn framework [5]. FreMEn is a mathematical tool based on the Fourier Transform, which considers the probability of a given state as a function of time and represents it by a combination of harmonic components. The idea is to treat a measured state as a signal, decompose it using the Fourier Transform, and obtain a frequency spectrum with the corresponding amplitudes, frequencies and phase shifts. Then, transferring the most prominent spectral components to the time domain provides an analytic expression representing the likelihood of that state at a given time in the past or future.

This model assumes that it is provided with people detection data, comprising the person's position, orientation and timestamp of the detection (x, y, α, t). When building the model, the x, y positions are discretised and assigned to the corresponding cell, and the orientation α is assigned to one of the k bins, whose value is incremented by 1. After a predefined interval of time, the bin values are normalised, and the results are used to update the spectra of the temporal models. Then, the bin values are reset to 0, and the counting starts again.

In order to retrieve the behaviour of human movement through a given cell at

a certain time t (which can be in the future or the past), the likelihood for each discretised orientation associated with a cell can be computed as:

$$p_\theta(t) = p_0 + \sum_{j=1}^{m} p_j cos(\omega_j t + \varphi_j), \tag{4.21}$$

where p_0 is the stationary probability, m is the number of the most prominent spectral components, and p_j, ω_j and φ_j are their amplitudes, periods and phases. The spectral components ω_j are drawn from a set of ω_s that covers periodicities ranging from 14 h to 1 week with the following distribution:

$$\omega_s = \frac{3600 \cdot 24 \cdot 7}{s}, \quad s \in 1, 2, 3, ..., 12. \tag{4.22}$$

Directional grid maps Directional grid maps (DGM) [92] are designed to model the *directional uncertainty* of dynamic environments. The inputs to the model are directions of objects at different locations of the environment, and the outputs are a set of continuous probability density functions indicating the most probable directions dynamic objects move at various locations of the environment. In order to build a DGM, firstly, the 2D or 3D environment is divided into a fixed-sized grid. Then, a mixture of von Mises distribution is assigned to each cell to model the multimodal angular uncertainty. Analogous to a Gaussian distribution, however, with a limited $[-\pi, +\pi]$ support, a von Mises distribution is controlled by its mean angle and concentration (inverse variance) parameters. The number of von Mises components for each mixture is determined by the number of density-wide clusters using the DBSCAN algorithm. Having initialised the von Mises distributions with the cluster centres, the parameters are learned using Expectation-Maximization (EM). In experiments, it takes 1 to 4 iterations to converge the EM. Since the directional grid maps are not designed to deal with the spatiotemporal domain with periodic patterns, in this experiment, the temporal domain is also discretised every 15 minutes in addition to the 2 m $\times$ 2 m spatial discretisation. For this experiment, as a proxy, we attempt to estimate the people density by considering the cells where the initial set of mixture parameters changes with time. Therefore, the proxy count probabilities are always either 0 or 1 and not the exact people density. In the future, it is possible to replace the von

Mises distribution with a Gaussian distribution or replace the Bernoulli likelihood in [52] with a Gaussian likelihood to model such spatial density estimations accurately.

CLiFF-Map Model Circular Linear Flow Field map (CLiFF-map) [73] is a technique for encoding movement patterns as a field of Gaussian mixtures. They can be combined with semi-wrapped Gaussian mixture models (SWGMM) to model multimodal motion.

$$p(\boldsymbol{V}|\boldsymbol{\xi}) = \sum_{j=1}^{J} \pi_j \mathcal{N}^{\mathcal{SW}}_{\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j}(\boldsymbol{V}) \tag{4.23}$$

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

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

$$\mathcal{N}^{\mathcal{SW}}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{V}) = \sum_{k \in \mathbf{Z}} \mathcal{N}_{\boldsymbol{\mu},\boldsymbol{\Sigma}} \left(\begin{bmatrix} \theta \\ \rho \end{bmatrix} + 2\pi \begin{bmatrix} k \\ 0 \end{bmatrix} \right). \tag{4.24}$$

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

LSTM We also implemented a deep-learning model for a point of comparison. A long short-term memory [93] neural network was built using Keras atop the TensorFlow library. It consisted of 4 layers of 50 LSTM units followed by a fully connected layer with 72000 trainable parameters.

Results

The evaluation results, summarised in Table 4.3, indicate that HyTS achieved the lowest root mean squared error, but for the χ^2 distance, STeF presents the lower score. We believe that the reason behind that is because STeF represents each orientation in each position by its own temporal model. Although DGM is not designed to estimate the probability of occurrence and only returns 0 or 1, and

Table 4.2: Qualitative Comparison Of Methods

Method		Time		Representation					Complexity	
Name	References	long-term	short-term	time	space	intensity	direction	speed	Memory [kB]	Train time [s]
HyTS	[19]	✓	×	C	C	C	C	C	**2**	60
STeF	[64]	✓	×	C	D	×	D	×	140	**20**
DGM	[92]	✓	×	D	D	C	C	×	20	72
CLiFF	[73]	×	×	×	D	C	C	C	6k	10^4
LSTM	[93]	×	✓	C	C	C	C	C	900	10^6

Note 1: In the 'Representation' columns, C stands for the continuous, and D is for the discrete representation of variables the method provides.
Note 2: CLiFF-map was developed using Matlab; other methods are written in Python.

therefore RMSE is not a proper measurement for this method, it was comparable with HyTS in χ^2 statistics.

To fit the evaluation procedure, CLiFF-map was discretised into eight orientation bins. CLiFF-map predicts directions on specific positions better than HyTS during the day, but as this model does not consider the temporal dimension, its night results are worse.

Low values of RMSE in Table 4.3 show that HyTS can accurately model both movement directions and human occurrences. Modelling the joint probability of occurrences and directions is the crucial property of the HyTS, which is supposed to provide apriori knowledge of the dynamics of human-populated environments to autonomous robots.

We also include an LSTM model, commonly considered a state-of-the-art method for temporal predictions. The LSTM model was trained using four NVIDIA Tesla V100 SXM2 32GB for 10 hours, and its model size was 900 KiB. However, as the LSTM is tailored for short-term predictions (compared to the prediction horizons of STeF or HyTS), its predictions quickly converge to the mean probability of people's directions across the entire training dataset (spatially and temporally). Therefore, LSTM predicts that the distribution of people's move-

Table 4.3: Prediction errors of the evaluated models and datasets

Testing sets	*Days*		*Nights*		*Days and nights*	
Criterion	*RMSE*	χ^2	*RMSE*	χ^2	*RMSE*	χ^2
HyTS	**0.50**	23.4	**0.00**	0.2	**0.40**	23.6
STeF	0.57	**10.6**	0.02	8.1	0.46	**18.7**
DGM	0.70	25.5	0.83	**0.0**	0.75	25.5
CLiFF	0.60	15.5	0.16	9.2	0.50	24.7
LSTM	0.57	25.5	0.22	**0.0**	0.48	25.5

ment directions will be uniform in the long-term horizon. DGM and HyTS predict very low probabilities of people's presence during the night, which corresponds to uniform distribution of walking directions. Thus, the χ^2 metric for these three methods for the night data is lower than STeF and CLiFF.

HyTS, STeF and DGM were developed in Python language, their training on regular personal notebooks lasted about one minute, and the model sizes are 2 KiB, 140 KiB, and 20 KiB respectively, which indicates that they could be applied in a field robotic tasks. It should be noted that the model created by HyTS is smaller by magnitude(s) than its competitors, which is an essential attribute for building models over large areas, see Table 4.2.

5 Summary

We developed a continuous version of FreMEn denoted as HyTS. We apply HyTS to different datasets, representing people's presence over several weeks. The evaluations demonstrated that the proposed representation allows for capturing people's habits. Their model can predict future people's presence and flows across corridors and identify unusual occurrences of persons within a given area.

The proposed method successfully competed with or outperformed state-of-the-art methods in chosen scenarios. It can be used in a wide variety of autonomous robotic tasks while the size of a model is significantly smaller than that of its competitors. It is comparable in computational time with the most efficient state-of-the-art methods. These attributes indicate its applicability in long-term autonomous robotic systems.

FreMEn and later HyTS showed that their ability to predict the future structure of an environment improves robotic mapping [5], [21], [55], [63], localisation [5], [15], [16], [94], path planning [17], [23], [60], exploration [10], [12], [64], [66], [95], task scheduling [17], [23], [67], patrolling [66], [96], searching [61], novelty detection [5], [18], [55], [97], activity recognition [98], human-robot interaction [17], [62], [65], [68], demand forecasting [26], and helped people avoid crowded places during the COVID19 outbreak in Czech Republic [99], [100].

Part III

Methodology for Benchmarking Long-Term Spatio-Temporal Maps

6 Outline

The last decade showed that the advances in robotics enable autonomous robots to operate in human-populated environments [9], [101], [102]. Such the environment includes diverse types of dynamics, such as natural daily [55], and seasonal changes [103], but most importantly, the dynamics imposed by human actions [3]. The robots that consider these dynamics and adjust their decisions provide better performance of their human-centric tasks [9]–[13], [104]. Moreover, later works assert that decisions of autonomous robots that consider human habits are crucial for robots to be accepted by human society [23], [74], [101]. We claim that robots, supposed to be part of human society, need to adopt the temporal structure derived from human customs to avoid polluting the psychosocial atmosphere and achieve long-term acceptance of their services by people.

We state that long-term human-aware navigation needs to include spatio-temporal maps in the navigation system to support the decisions before the navigational task. The traditional approach to robot navigation in an uncontrolled environment combines static maps [28], [71] with a reactive approach to unexpected events. However, many unexpected events are usually caused by human actions, especially human movement through the robot's operational environment. The reactive replanning of a robot's trajectory based on sense-plan-act frameworks, used, for example, in Robot Operating System [105] move base, gives people an impression of clumsiness [106] due to their slow response, which eventually leads to negative emotions towards the robot [23]. This unwanted interaction is mainly driven by the fact that by the time the robot can detect people walking around it and replan its trajectories, the person is already taking evasive action. Therefore, incorporating the expected human movement through the environment into the robot's navigation and planning is of high importance for the autonomous robots

intended to help people in their environment [107]–[109]. To show the fundamental impact on human acceptance when incorporating spatio-temporal maps into navigation, we provide a field robot experiment with an industrial-grade robotic platform. The experiment studied how many people traversing the university hall were irritated by the robot planning its navigation with and without the map.

As the spatio-temporal maps were used in different scientific fields, we created a comprehensive overview of the spatio-temporal representations. They provide various aggregations, like frequency, likelihood, probability, or relative weights. The authors compare the methods in different ways, on different data, and using different metrics, making it difficult to assess their actual performance. To overcome the difficulties in comparison of different approaches to mapping, we proposed the methodology [17] to create different criteria from a chosen cost function that suits the studied attributes of maps. We created an extensive set of generalised and simplified approaches to represent the spatio-temporal phenomena covering a large part of the current methods. We applied them to the real-world dataset of one month of human detections and compared them using criteria following the proposed methodology. The straightforward interpretability of the results provides us with exceptional insight into the strengths and weaknesses of different approaches.

7 Related Work

7.1 Static Maps

7.1.1 Discrete Maps

A straightforward way to include a historical knowledge of human behaviour into the robotic map was proposed in [110]. The authors divided the map into the crossings (nodes) and halls (vertices) and projected the map into the graph. A set of cameras scanned every hall, and the visual system detected and counted people in these halls. Each vertex of the graph created a dictionary of observed human movements, which leads to a historical heat map. The field robot experiment proved that navigation using historical knowledge leads to evading congestions. The authors of [111] included in the classical grid-based map historical information about the frequency of people in every cell, usual directions and speeds and also acceleration. They could detect natural ways people use in larger areas, where the usual paths are not apparent from the hall's structure. They stated that the knowledge of the acceleration distribution leads to a better understanding of "smooth flows" in a human movement. In [92] the authors proposed a Directional grid map (DGM) that models the distribution of directions and speeds in the occupancy grid. Each cell applies an expectation-maximisation algorithm to estimate the mixture of von Misses distributions from historical data.

7.1.2 Spatially Continuous Maps

The authors of [27] argued that a continuous spatial map has better properties for robot navigation than a classical grid map. They proposed Gaussian process

occupancy maps (GPMOs) that overcome single-scaled maps' problems and provide accurate maps even with relatively sparse and noisy data. However, such an approach cannot be applied to modelling directions. It suffers from averaging angles, leading to meaningless predictions when humans traverse positions in both directions [112]. As a solution, they proposed a combination of the Kernel Bayes' Rule and the Gaussian mixture pre-image recovery method (KBR-GM). Another approach to overcoming weaknesses of discrete occupancy maps was proposed in [113]. The authors did not use Gaussian processes mainly because they generally fail to model multi-modal distributions. Instead, they modelled the continuous distribution of measured directions and speeds in every cell of a grid using an expectation-maximisation method based on the Independent von Mises–Gaussian distribution [114]. Then they defined an interpolation technique that provides a distribution estimation at any position. They called it CLiFF-map and applied that to the people tracking and wind data. In the following work [73] they trained the CLiFF-map with sparse data (75% and 3%) and then reconstructed it to the original resolution using Monte Carlo and Nadaraya Watson methods. They proved that it is possible to gather relatively accurate models from very sparse data. Later, they propose a robot motion planner over the CLiFF-map based on RRT* [74]. They define the Extended Upstream Criterion (EUC) as a cost function to favour the robot's movement with the vector field. The trajectory found in such a way is most likely to offer a good tradeoff between length and control effort against the dynamics of the environment. They showed that the model of the movement of people could be used to navigate the robot through the crowd effectively. In [115] CLiFF-map was extended with a Down-The-CLiFF (DTC) cost function for trajectory planning, which explicitly accounts for the environment's dynamics and the uncertainty in the flow model. The cost function incorporates observation and motion ratios, which, compared to the previous EUC, favours the less crowded areas and provides information for an exploration task.

7.2 Short Term Dynamics in Maps

In the aforementioned approaches, historical knowledge does not provide any information about the changes in the dynamics of an environment. One of the approaches is to include the short-term changes in the map derived from the actual observation of the situation. In [4], the authors proposed a Conditional Transition Map (CTMap), which models in every cell the probability distribution of vehicle transitions between the last and next cell. As the behaviour of the usual vehicle is not random, the CTMap estimates the probability of the exit direction of the vehicle from the entry direction. Using the Conditional Probability Propagation Tree (CPPTree) graph representing all reachable transitions, the method can predict the vehicle's trajectory that entered the scene or provide a convenient navigational plan for a robot. Another discrete short-term model can be found in [72], where authors predict the path of a walking person based on his actual position in the environment using the input-output Markov model. The model that estimates a human's future position directly from observations [116] allows for reactive navigation that respects a human's personal space without the need of repetitive recalculations of robot's trajectory. [117] proposed a static spatial map that includes distributions of directions and speeds of the vehicles at the crossroads. It is used as prior to the trajectory prediction, which is updated based on the current observation. The current observation, for example, the blinking of a car, is then used to estimate the new directional distribution and eventually the most probable path of the vehicle. The authors of [118] propose a graph that represents an environment and which edges were traversable or not during a robot's task. The model learns which edges were traversable together. Based on this model, they can predict where the robot can go after observing a small part of the environment. Then, the information is exploited to provide better navigation plans. With the advancement of the neural networks that include long short-term memory, one can find in the literature a vast amount of the methods for short-term predictions derived from the actual observation [107].

7.3 Partly Discrete Spatio-temporal Maps

7.3.1 Maps with Discrete Temporal Domain

Apart from modelling the directions of the movement of particles in the environment, the scientific community focuses on modelling the changes in the environment over time. A prevalent approach to model a time-dependent phenomenon is to create seasonal windows (usually one day long) and to model the environment over the different parts of these windows [37], [38]. This approach models patterns of the environment change, but it is necessary to define the principal periodicity, i.e., the length and the resolutions of windows. In [40] the authors analyse the flow of the wind. They proposed a continuous AirFlow Map and called it stf-AFM. It is a set of spatio-(short)temporal models inserted into the "calendar". The authors are searching for periodical features over the calendar using a combination of autocorrelation, fast Fourier transform and clustering over the frequencies. They pointed out that "knowing periodicities could have some impact". The bachelor thesis [119] proposed a method of Time-window GMM applied to the taxi demand prediction. Contrary to the previous article, the essential periods were gathered from the data in advance and used for proper time windows creation. The method then estimates the spatial distribution using EM-GMM in every time window. The author showed that it is possible to predict the demand for a few weeks to the future. A comparable method was also used for ambulance calls forecasting [41]. The authors proposed a spatio-temporal log-Gaussian Cox process. They divide the prediction task into multiple subtasks for different temporal and spatial classes (7 days of the week, 4 seasons of the year, and 5 regions). In those 140 subsets, they calculate a continuous one-day-long model derived from historical data of several years.

7.3.2 Smoothing the Grid

However, the time intervals usage suffers from the discontinuity at the borders of the intervals, which is necessary to overcome [75]. The bachelor thesis [120] describes the system that was created to help people during the early times of

the covid pandemic. The system predicts how crowded will be places like shops and pharmacies [100]. As the learning data were highly sparse and unevenly measured, the author employed a histogram-based temporal model smoothed by spline at every tested place. Despite the meagre quality data, the system could predict the ideal time for shopping in the next few days. The smoothing of the spatio-temporal models is generally a complex task. For example, the authors of [121] applied triangulation over the space of a discrete spatio-temporal model and interpolated the distribution linearly over the sub-spaces so that the distribution is piece-wise linear and continuous.

7.3.3 Continuous Modelling of Time over the Spatial Grid

A specific approach to spatio-temporal modelling is called Frequency Map Enhancement (FreMEn) [5]. In one of the experiments, the authors created a topological map. Each place on the map is defined by a set of visual features detected in the past. Every detection gets an ID, and if detected multiple times, it creates a binary time series (detected during observation or not). Every time series is then decomposed by frequency analysis. Each place at any time is then described by a set of the most probably detectable features. It is also possible to represent each place as an occupancy grid where each cell holds the binary time series that specify the occupancy state. All modelled phenomena are understood as static or periodic, which allows predicting future states of any feature at any time. Therefore, it is also possible to predict the most probable position of static and dynamic obstacles. The authors also define the persistence of the features, which allows the map to include unpredicted obstacles that were actually detected. Moreover, an open-source library integrated with ROS for long-term mobile robot mapping called FROctomap [57] is available.

Experiments and applications of FreMEn provide the most complex insight into benefits obtainable from incorporating spatio-temporal models into robotics. FreMEn was applied in the task of mapping [55]. The experiments showed the ability of the method to model binary states with very high precision. Applying it to the occupancy grid map lowered the prediction error compared to the static approach by 60%. Then, it was applied to the localisation task [15]. The experiment

was accomplished at eight places that were supposed to be distinguished using two different approaches. The first approach described each place by a "fremenized" occupancy grid, and the second approach modelled the places by image features whose detection probabilities were also fremenized. The experiment proved that incorporating the FreMEn improved the localisation ability of the robot. They also stated that the most prominent period was most influential in a long-term experiment as it persists over a more extended time. The authors of [60] proposed a topological map in which was the traversability of the edges modelled by FreMEn. They implemented the time-indexed Navigation Markov decision process that improved planning the navigational tasks in the changing environment. In a robotic search task [14], the performance of Frequency Map Enhancement based and Periodic Gaussian Mixture based modelling were compared. The robot was supposed to search for people and objects in three different environments. Compared to the stationary models, the experiment showed a decrease in the search time of about 25% and a decrease of places of about 33% by both proposed methods. The authors speculate that finding periods using FreMEn and approximating the events with the mixture of Gaussian distributions can lead to a better output. Although the fremenized occupancy grid holds the parameters of the binary time series in every cell, it was successfully applied to a directional grid map over a large area of mall halls [63]. FreMEn was also subsumed into the *Human-Aware Allocation* proposed for cooperation between multiple robots in a human-populated area [65]. Its predictions were employed in the multilayer *Map of Dynamics* to ensure the optimal division of tasks with human presence in the environment. Considering the ability of a robot that uses FreMEn to recognise "when" it observed the most informative situation, the novel *information-based Monte-Carlo scheduler* for exploration was proposed [10]. Eventualy, different exploration strategies [12], [64], [66] and exploration-exploitation dilema [67] were studied. In [62], FreMEn was redefined into the Addition Amplitude Model (AAM). The most significant frequencies were iteratively discovered from learning data and actual reconstruction differences, i.e., model errors. The main difference to the original FreMEn resides when the frequency with the highest amplitude in the errors is part of the frequencies that established the reconstruction. In such a case, the amplitudes of those two (identical) frequencies are summed up, and the shift is averaged.

7.4 Continuous Spatio-temporal Maps

The main problem of the approach above is the spatial independence of the neighbouring cells. It can be addressed by spatial ordering [122], but the spatial ordering has to be predefined a priori and cannot be changed during the learning process [123]. The necessity to model spatial dependencies in the dynamic environment leads to a continuous spatio-temporal representation of the environment. Although the continuous models of the environment changes are computationally demanding, they are beneficial due to their memory efficiency[20], [27] and robustness to the outliers [124].

7.4.1 Maps with Predefined Periods

The continuous spatio-temporal models were successfully used in ambulance demand prediction. The authors of [39] proposed a time-varying Gaussian mixture model. They created two-hour windows with a one-week periodicity, and in each window applied GMM. Moreover, they apply constraints for the weights of each Gaussian in a way that timely preceding and subsequent weights, and weights from identical time windows of preceding and subsequent days, directly influence the current weight. In this way, they include into their model also one-day periodicity and smoothness to the model change, which they refer to as *short-term serial dependence*. Later, they proposed a warped kernel density estimation model that produces geometrically better density estimation than time-varying GMM [43]. The warping kernels are derived from spatial geometry. The spatial positions of the ambulance demand constitute clusters projected to a weighted graph as nodes. The edges are derived from the highways and roads, and weights correspond to the usual traffic. The parameters of the graph then influence the shapes of the spatial clusters. Such an approach provides a "fit to structure" distribution. The fundamental periodicity in the proposed model was one week. Due to the computational complexity, the model used an 8 weeks-long sliding window for learning while prediction was tested in consequent 4 weeks. Continual learning also served as "forgetting" and provided different predictions in different seasons (winter/summer). However, the model did not prove its improvement over initial

methods when applied to a crime prediction in Bogota [45].

In the following work [46], kernel warping was applied to the homicide prediction in Bogota. The authors explain the inefficiency of the model in several ways: the relatively low amount of homicides, changes of positions of problematic groups across historical data, and the direct impact of patrolling strategy of police known as "runaway feedback loops". They solve the known issues by enriching the data with street fights, which are very connected to the homicides, and including temporal decay components in the studied models. Temporal decay helped all the models, and the kernel warping model provided the best predictions. In [47], the spatio-temporal ambulance demand is modelled in a way that the dataset is divided into spatio-temporal cells. By applying autocorrelation to the data, the authors estimated two main periods. Then those periods were exploited to create a continuous spatio-temporal model by applying kernel-based GMM to the weighted aggregations in the cells. The weights were derived from the distances of cells in time considering found periods. The method is referred to as spatio-temporal kernel density estimation (stKDE). stKDE was also used for a modelling ambulance intervention in Milan [48]. The authors pointed out that it is possible to use fast Fourier transformation instead of autocorrelation to estimate periods in the data. In [49], a spatio-temporal model based on stKDE with one periodicity was used as a part of the method for an allocation of ambulance bases.

Spatio-temporal models were also applied to the spread of disease prediction. In [42], the authors modelled the propagation of influenza. They created a spatio-temporal model based on Gaussian processes consisting of spatial, temporal and spatio-temporal components. Except for the temporal decay, the model also consisted of the one-year periodical temporal component. The periodicity was chosen based on general knowledge. Similarly, a continuous spatio-temporal map of disease spread over Turkey consisted of one year periodicity [44]. The authors stated that the continuous spatio-temporal model could predict the disease spread in unmeasured regions inside Turkey in the future.

7.4.2 Periods Retrieved from the Data

Although the aforementioned models were continuous, spatio-temporal, and included seasonality, only one or two periodicities were employed. In many cases, the periodical component was obtained not from data but a priori knowledge or expertise. In long-term robotics, a robot has to obtain the natural periodicities of the dynamic environment from its observations, and some of them are likely out of usual human expertise. Moreover, some human-populated environments do not necessarily follow the day/week/year pattern. The flow of continuous media was modelled in [50] by applying Gaussian processes to the measured data. The algorithm used a covariance matrix formed by the spatial and temporal components. The periodical temporal components were obtained iteratively using frequency analysis over the model's errors. The temporal components also include temporal decay. The spatial and temporal components of the covariance function are calculated separately and multiplied at the end. Nevertheless, the computational complexity of the Gaussian processes grows fast with the number of data points [62], [125]. The extraction of multiple periodical features from the data was targeted in [53] proposing Fourier Feature Approximations for Periodic Kernels and its multidimensional variant [54]. Although it was not applied to the spatio-temporal modelling, together with formerly proposed Hilbert maps [51] that were subsequently improved to be updated incrementally [52], it should be taken into consideration.

In [16], the authors proposed a continuous spatio-temporal model based on a projection of linear time into the closed subset of higher dimensional vector space, the Warped Hypertime. The projection has its roots in the seasonal windows, but the windows are coiled into circles. Such an approach ensure continuity of the model on the edges of a window. Projected circles form a multidimensional Clifford hyper-torus [126]. The seasonality is obtained iteratively from an error of the actual model. The error forms a time series that is analysed by FreMEn. The FreMEn returns the most dominant periodicity found in error, and the periodicity forms a new circle. The authors then showed in experiments that the proposed method is equal to or better than FreMEn when applied to the robotics tasks. In [18], Warped Hypertime was applied to the task of detecting anomalies. It

was proved that the model learning on binary data is faster and converges to a better quality model than other state-of-the-art methods. Later, it was applied to the detection of novelties [97] in one-dimensional time series without any trend, and it proved the similar quality of the model to the state-of-the-art regression methods while performing better in modelling time series with multimodality. Applied to a human presence prediction [20], it showed high memory efficiency with a slight boost to the prediction compared to the discrete methods. Authors of [127] applied a similar approach, but they use Gaussian mixture models with periodic kernel for modelling the pedestrian frequencies in the ATC dataseti [128]. The data for the training were gathered by a simulation of the robot's movement through the environment. Only data in the simulated field of view of the robot were taken into account. Experiments proved the ability of their continuous spatio-temporal method to create a useful map from very sparse data. In [21], the hypertime version derived from the binary version was applied to the directions of people. To avoid ambiguity caused by averaging angles, the model's spatial part was four-dimensional - the method also modelled speeds, i.e., velocities, and the resulting angles were calculated by integrating velocities over the angular intervals. It was not computationally rational to calibrate the model using a grid of aggregations similar to the previous work [21]. Instead, the authors extended the data with "negative measurements"; they added random noise labelled as "not a human". Two models (humans and artificial nonhumans) were learned, and the "calibration" was done similarly to the approach used for binary data [18]. Although the addition of artificial zeros is meaningful in some two-dimensional spatial cases [27], generally, it is not a convenient approach - weigh how many possibilities of "not detected velocity" one has to take into consideration.

7.5 Benchmarking Spatio-Temporal Maps

The authors that propose new mapping methods apply a wide variety of quality measuring techniques. It is not uncommon to provide only a discussion about the visual quality of the map regarding the most common directions [4], [113], reconstructed signal [60], or changes of heat map over time [49], [76]. Such an approach is usually used to provide insight into the proposed concept's basic

behaviour.

Very popular measure of the quality of discrete or discretized maps is mean square error [16], [42], [83], [92] and rooted mean square error [21], [43], [50], [62], [117]. Apart from the fact that, in general, the maps not providing the frequency of measurements need to be normalised [44], the difference between maps is very small [21], [23] which leads to the necessity to enhance the differences of results [20], and the rank of methods is dependent on the coarseness of the grid [119]. Similarly, a popular measure is an average (negative) log-likelihood that can be applied to the continuous maps [43], [47], [92], [112], [124] but is only meaningful when a probability measure can be derived. It suffers mainly because the map is tested only against detections, i.e. in places with testing data points.

Many authors treat their maps as binary predictors and test their quality using measurements derived from the confusion matrix. For the predictors with predefined threshold, we can find a ratio of true positives [55], [129], ratio of false positives [63], hit rate [45], [46], and accuracy [130]. Some authors do not predefine the thresholds and use measures like an area under the curve [27].

We can find also unique measurements of a map quality like Pearson correlation coefficient [44], Cramer-von-Mises criterion [40], Kullback-Leibler divergence [107], and k-NN Universal Divergence Estimator [73]. Rarely, we can find average probability density [92], [117], Chi-square distance [21], [64], [127].

In the robotic community, there exists also possibility to compare the maps in simulations [74], [118], or simulated robotic tasks built on the real data [5], [14], [15], [23]. The ultimately self-evident measures of the quality of the map then come with real robot experiments [5], [110].

8 Generalised Natural Criterion for Benchmarking Spatio-Temporal Maps

8.1 Original Idea

Previous work highlighted the question of the proper evaluation of spatio-temporal models for service robotics in previous work [21], [23]. Some authors strongly imply that generic criteria used to quantify the success of regression methods, for example, mean squared error (MSE), are unsuitable for this task [21], [119], [131]. It is mentioned that these criteria struggle to differentiate between state-of-the-art methods meaningfully and are not reliable. Small changes in the hyperparameters of the testing procedures can cause significantly different results, even changing the rank of the methods. More importantly, they do not provide a useful proxy of the measure of how well the people in a robot's environment perceive the robot's behaviour, i.e. some *robot acceptance* cost (RA), as they generally have no connection to the application [131].

This work is built upon a criterion called *expected encounters* (EE) developed in [23]. We considered it a natural criterion for comparing pedestrian flow forecasting models. However, spatio-temporal maps could and should be evaluated from various perspectives [17]. Therefore, we propose a generalised definition leading to a spatio-temporal maps' benchmarking methodology. To make the text uncluttered, we first define EE according to [23], and then we generalise the ideas to get a universal framework for defining similar metrics for evaluating

various robotic maps. Afterwards, we define the specific criteria used for the methods' comparison.

8.2 Expected Encounters

EE falls into the class of utility metrics, and its primary goal is to measure the usefulness of the learned model for the robot's navigation in human-populated environments. The EE criterion specifically tests the method's ability to model the phenomena of human presence in time and space. It does so by presenting a robot with a trained model and a set of navigational tasks that happen at different times. The robot is then supposed to plan its path through the environment for all these tasks minimising the number of people it expects would cross its path. Path planning tests the method's ability to model the phenomena in the spatial domain. Another principal idea is introduced to test the temporal domain: the *servicing ratio*. The robot is asked to select a specific percentage of times out of all the times presented where it will execute the navigational task, again to minimise the expected number of encounters with people. All the selected plans are then executed and compared with the positions and movement of people in the testing dataset. The testing environment calculates the encounters between the 'blind' robot and 'blind' people—the system does not provide any artificial reaction of any element involved.

To formally define the original EE criterion, we first define the *service disturbance function* $SD(r)$ with the parameter $r \in [0, 1]$, which we interpret as the *servicing ratio*, i.e. the ratio of navigational tasks the robot is required to perform. The robot is given a set of navigational tasks at times $\{t_i\}_{i=1}^{N}$, and for each one of them, it plans a path with cost c_i defined by the spatio-temporal model of human presence to be evaluated. The testing dataset then, for a given trajectory, gives the number of weighted encounters e_i (see Section 9.2.1) that would occur for the given planned trajectory. The *service disturbance function* is then given as

$$SD(r) = \sum_{k=1}^{\lfloor rN \rfloor} e_{\pi(k)}, \tag{8.1}$$

where π is a permutation which orders times t_i so that $\forall k : c_{\pi(k)} \leq c_{\pi(k+1)}$, where

c_i corresponds to t_i, as defined above.

The interpretability of the *service disturbance function* alone provides us with quite an interesting insight into the performance of individual models. To arrive at a single-value aggregation of this function, we adopt the interpretation where we understand this function to be the quantile function of the ordered human-robot encounters. The quantile function allows us to compute the expected value of the distribution, which we denote *expected encounters*, and we compute it as:

$$EE = \int_0^1 SD(r)dr. \tag{8.2}$$

Note that the service disturbance function satisfies the necessary conditions of being cumulative and defined on the $[0, 1]$ interval.

8.3 Generalised Definition

The criterion was explicitly developed for comparing different approaches to the spatio-temporal modelling of pedestrian flows [21], [23] with the necessity of evaluating the predictions producing very different values. Consider the outputs of histograms over seasonal windows [38], CLiFF-map using different cost functions [74], [115], Time-window GMM [119] producing the values limited only from below by zero, binary-map approaches based on Frequency Map Enhancement [5], and probabilistic STeF-map [64]). However, the original criterion focuses only on one specific aspect of human-aware navigation. There are a lot of different opinions as to what criterion influences the acceptability of a robot in human society most [132], [133]. Besides, the acceptability need not necessarily be the only point of view that defines the quality of performing the task [14], [60], [67].

The general idea of comparing the spatio-temporal maps is applicable and helpful in comparing different robotic models predicting diverse phenomena useful in miscellaneous robotic tasks. There are two essential prerequisites needed for the application of the idea of the criterion—a predictor producing (at least) ordinal values that are tailored for the robot to decide on its task, and a measure of the impact of the robot's decision usually denoted as a *cost*.

To formalise the generalised definition, we start by having a specific task for

the robot to perform, during which it relies on information provided by a model and an observable measure of how expensive it was for the robot to perform its task—the *true cost*. This true cost can but does not have to be directly tied to the predictions of the model—a *predicted cost*. This task is supposed to be performed under different values of the model's explanatory variable (e.g. time or space) $\{v_i\}_{i=1}^{N}$ and for every i the robot acquires a predicted cost $c(v_i)$ of its hypothetical task from the predictor. Then, we define a *true cost function* $\mathcal{C} : \{1, \ldots, N\} \to \mathbb{R}$, which for every i provides a true cost. Finally, we define a *quantile function* of the distribution of the ordered true costs as:

$$Q(r) = \sum_{k=1}^{\lfloor rN \rfloor} \mathcal{C}(\pi(k)), \tag{8.3}$$

where π is a permutation, which orders indexes i so that $\forall k : c(v_{\pi(k)}) \leq c(v_{\pi(k+1)})$ All interpretations like the *servicing ratio* hold in the generalised case as well and it is again possible to compute the *expected value* of the ordered true costs:

$$E\mathcal{C} = \int_0^1 Q(r)dr. \tag{8.4}$$

8.4 Methodology

We strongly disagree with applying statistical tools developed to describe or compare repeatable experiments to experiments involving people. The same applies to other uncontrollable entities, like animals, wind, or traffic. For human-centred robotics, we need to follow the general idea of natural experiments [134] and apply the methods in real-world situations. However, autonomous robotics is not in the development stage in which the field robot experiments can be accomplished securely in large enough repetitions to provide comparable results. We suggest that the comparison should be performed (at least) in the simulations built on the real-world data. In these simulations, we can repeat the same real situation and apply the different methods.

The proposed generalised definition of the criterion can be applied to various robotic tasks a robot performs. We need to specify the predicted costs $c(v_i)$ and

the true cost function $\mathcal{C}$ that provides the true cost during the robot employment in the simulated, real-data based experiment. Using the proper definition of the true cost function, we can compare the approaches and specify which aspect of the proposed approach is better than another. Moreover, we can also elaborate on what aspects should be compared by designing new criteria following the generalised one. As the proposed criterion and its usage is derived from the general idea of natural experiments, we denote it the *Generalised Natural Criterion*, *GNC*.

The general template for applying GNC for a particular predictor then has the following steps:

1. execute or simulate the experiment with different settings of the investigated environment,
2. order the measured costs by the predictions of the method and accumulate it, as described in Equation 8.3,
3. calculate the expected value of the cost of robot's behaviour, as in Equation 8.4.

8.5 Application of GNC

GNC and the concept described above are generally applicable to the scenarios where the robot can estimate what the cost of solving the task in different situations is. The predictor providing the estimation uses as input some explanatory variables, like a relative position of a robot to an obstacle, the time of task execution, or actual weather.

Then, we need to define a function, referred to as true cost function $\mathcal{C}$, that provides a measurement of the quality of the performed task. For example, in [133], the quality of human-aware navigation is measured by the mean distance between the moving human and robot, the time spent in areas associated with the personal zone, the human discomfort, and the total navigation time needed for the robot to reach the desired goal. Such true costs are directly applicable, and the different approaches to solving the problem can be compared. In the case of testing different spatio-temporal maps [23], the authors defined the true cost function as the cost of the optimal path.

We mentioned that the predictor needs to produce at least ordinal values. The optimal predictor produces real values with the probability of obtaining identical values close to zero. The following sections discuss some scenarios where the direct application of the methodology defined above is not straightforward because the predictions are not real values. It needs some additional effort to utilise the method entirely.

8.5.1 Binary States

Many environmental models in mobile robotics are composed of independent binary states, such as the presence or absence of people, the visibility of landmarks, and the traversability of certain areas, see [5]. As an example, let us consider the task of traversing an environment using its time-varying topological map as quickly as possible [60], [118]. Although the robotic methods usually provide more nuanced prediction than just 0 or 1, e.g. a probability of the state being 1, it happens that a confident predictor forecasts that the traversability in more than one situation will be 0 or 1. In such a case, the ordering needed in Equation 8.1 would have several different solutions. We can divide the predictions into successful (1) and unsuccessful (0) classes and understand the values of 0 and 1 as an ordinal variable. Inside those classes, the ordering can be done randomly, and the expected cost value can be calculated as an average of multiple random reorderings.

8.5.2 Ordinal Variables

When the target variable is ordinal, the ordering can be solved similar to the binary states by multiple random reordering. If available, the predicted values can also be weighted by the system's confidence or the probability of each class. Another approach consists of applying a regressor function to predict the targeted value, as in the bachelor thesis [97]. Both approaches exploiting the confidence provide a continuous output of the prediction, and therefore, it is possible to order the costs with ease.

9 Evaluation

9.1 Human Disturbance Experiment

The assessment of how the occurring people in the environment accept the robot's presence is not straightforward. In some robotic papers, we can see, for example, an effort of the scientists to learn a robot to evade accessing human's personal space [133] or breaking pair-wise social relations [135]. However, the survey on the experience of people with the long-term autonomous robot [106] shows that a major disappointment is caused by the robot's inability to adapt its behaviour to the general pattern of human activity and learn that certain paths can be blocked at given times [136]. Overall, the robot was expected to be intelligent and adaptive, rather than just repetitively performing the tasks. The lack of adaptation results in breaking ongoing social interactions because of a more critical task (e.g., battery charging) or awkward evasion manoeuvres during human-robot encounters in crowded areas. From our observation, the inability of the robot to adapt its activity to the temporal patterns of its workplace causes it to be perceived as unintelligent, awkward, annoying and useless. Eventually, as a robot provides a service nobody wants at the right time, people start to push it to stay away, turn it off, or otherwise treat it as an unwanted entity.

While a robot certainly needs to adequately react to the arising situations and avoid people in a socially-acceptable way, people perceive the robot better if it can schedule and plan its activities to avoid socially inappropriate situations altogether.

Our primary hypothesis states that the robot that can decide when to provide a service disturbs people less than a robot without such ability. To prove the hy-

pothesis, we run a robot in a human-populated environment using two types of navigation systems. The first one used a traditional occupancy grid containing static obstacles to plan the paths, which were traversed using an industrial-grade reactive navigation system provided by the manufacturer of the HSR robot used for the experiment – we call this *Reactive* navigation. The second one, named *Anticipative* navigation, also used the build-in navigation of the HSR robot. However, the paths and times of navigation were planned using a spatio-temporal-directional map of pedestrian flows.

We measured people's discomfort during the experiment by counting only the strong negative reactions toward the robot, which eventually led to an intentional search for a way to complain. A human that goes and complains was, in our interpretation, so distracted, that he changed his original plan and started to solve the unexpected situation - similarly to the strong negative reactions that led to pushing a robot out of society (as mentioned earlier).

9.1.1 Robotic Platform

As shown in Fig. 9.1, a Toyota HSR robot [137] was employed for experiments, which is equipped with an Xtion RGB-D camera (for obstacle avoidance only), a Hokuyo UST-20LX 2D LiDAR (for obstacle avoidance and global path planning), and other sensors that were not used in the experiment. The human-aware navigation system was deployed on a laptop with an Intel i7-7700HQ processor and 32GB RAM, wired to the robot and carried on its back. All the software was implemented into the Robot Operating System (ROS) [105] with high modularity using C++ and Python mixed coding, running in real-time on Linux Ubuntu 18.04 LTS (64-bit) and ROS Melodic.

9.1.2 Experimental Setup

The real-world experiments were conducted with the HSR robot on the afternoon of the 12th of December and the morning of the 13th of December 2019, and both in the same place, the hall of the UTBM building. The experiments were designed by researchers who do not work at UTBM, and there was no advance notification to anyone involved in the experiment. In each experiment, we allocated two

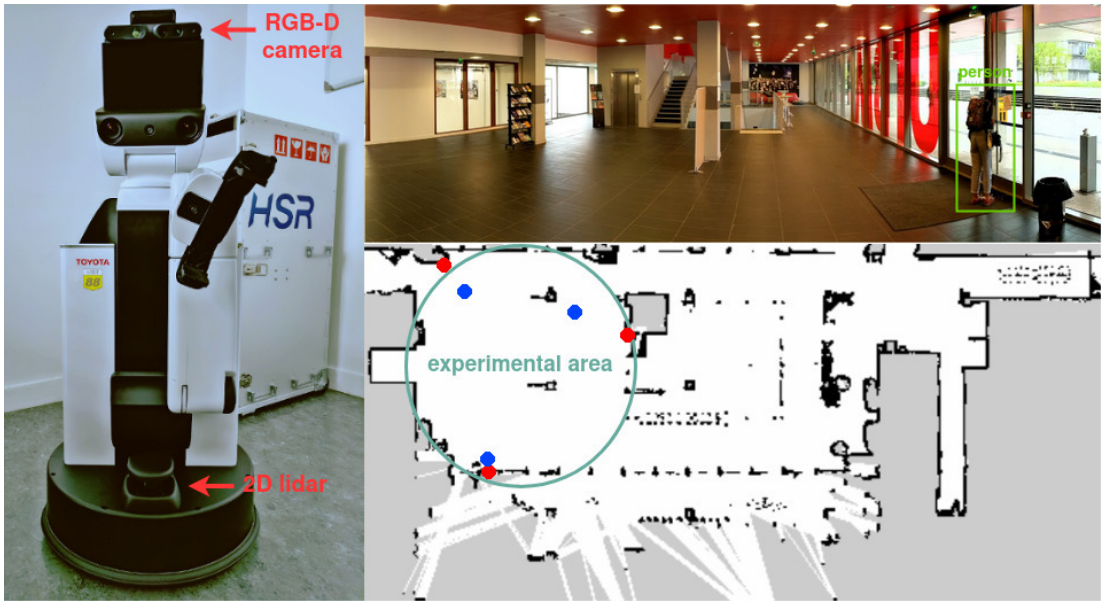

Figure 9.1: Left: Toyota HSR robot including an Xtion RGB-D camera and a Hokuyo UST-20LX 2D LiDAR. Upper right: experimental environment. Lower right: occupancy grid map for robot navigation. The red points are representative of the positions of the removable tags, and the blue points are the locations of the waypoints.

40 minute slots to perform 10 patrols, during which HSR had to visit three predefined waypoints. By integrating our spatio-temporal model with the HSR's own navigation system, the robot polled three waypoints in a counterclockwise sequence while adapting its movement to the moving people using HSR's built-in collision avoidance. If people occupy a waypoint, the robot waits until the waypoint is free.

Moreover, as shown in Figure 9.2, three paper sheets were placed with removable tags near the three waypoints. These sheets asked people to remove a tag if they felt that the robot was causing a nuisance by forcing them to avoid it. The idea was to count how many people were distracted by the robot that they performed an intentional operation due to the stressful situation.

9.1.3 Results of Human Disturbance Experiment

The robot was supposed to drive through the environment 10 times in every time slot. Every run took approximately 2 minutes, which means that the task

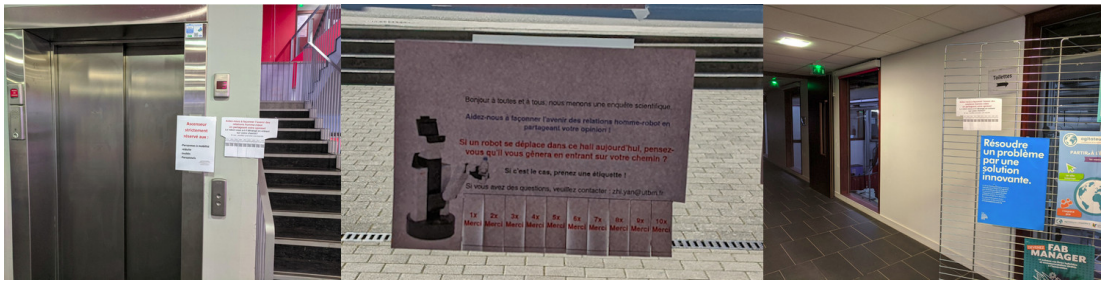

Figure 9.2: Three paper sheets placed with removable tags near the three waypoints.

Table 9.1: Comparison of reactive and anticipative navigation

Evaluated behaviour	Time	People total	People involved	People annoyed
Anticipative	9:20-10:00	115	17	0
Reactive	10:00-10:40	132	46	2
Anticipative	16:00-16:40	43	6	0
Reactive	16:40-17:20	23	14	1
Control	9:20-10:40	211	-	0

took up 50% of the assigned time. The reactive navigation went through the hall in uniformly distributed times, covering whole timeslots. The anticipative navigation decided on the proper times by exploiting the predictions from the spatio-temporal map of pedestrian flows (similar to FreMEn_WHyTeS_Clusters, Section 9.2.3). In Figure 9.3, the graph represents the prediction of the *robot acceptance* (RA) cost in the area, i.e., the disturbance to the people caused by hypothetical execution of the patrolling task regarding the prediction of pedestrian flows [23]. A *control* experiment was also performed to evaluate how often people would indicate they were annoyed with the robot regardless of its behaviour. The robot was removed from the vicinity of the experiment, but the recording apparatus was left in place.

The experimental results are summarized in Table 9.1. The column 'People total' describes the number of people passing through the area within the appropriate time window. The number of people in the hall while the robot was performing its task is listed in the 'People involved' column. The last column summarizes a number of paper tags removed by people that considered the robot's

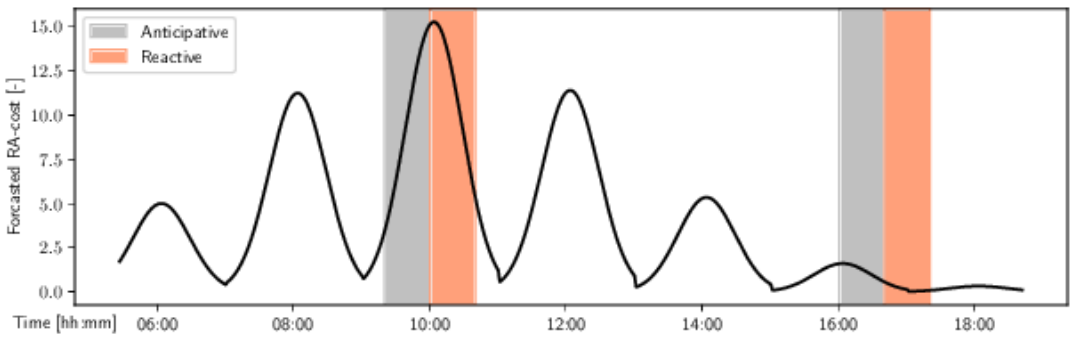

Figure 9.3: Forecasted (8 month horizon) robot acceptance (RA) cost and timeslots allocated for the experiment. Note that the bottom of the troughs are sharp due to the model not using sinusoids. Courtesy of [23].

behaviour annoying.

The morning timeslots chosen for the experiment included more people walking through the hall than evening timeslots. The observed distribution of people was not uniform. For example, when the robot was performing its tasks, we manually counted that in the time (sub)slots 9:50–10:00, 10:00–10:15, and 16:00–16:10, there were approximately 70, 90, and 20 people passing through, respectively. It corresponds to the prediction of the RA cost, Figure 9.3.

While performing the control experiment, 211 people passed through the hall, and none of them recorded that the robot was annoying them. That indicates that the people passing through the hall were not distracted by the equipment so that anybody would search for a way to complain. The robot taking advantage of anticipative navigation decided to perform all the tasks during time slots 9:20–9:40 and 16:20–16:40. Incorporating the spatio-temporal model into the robot's planning led to evading the most crowded time slots, as seen in the column 'People involved'. The robot incorporating only reactive navigation and performing its tasks, uniformly drove through the hall even in very crowded situations when people were in a hurry. The robot's presence inevitably affected the flow. People were distracted by replanning their trajectories, which resulted in the intentional change of their original plan from 'to pass the hall' to 'to complain' - as documented in the 'People annoyed' column.

We proved that the navigation system that follows human routines distracts

people less than the one lacking this capability. We also showed that predictions from the spatio-temporal maps of pedestrian flows provide good enough information to evade socially inappropriate behaviour and strongly support reactive navigation. The difference between the anticipative and reactive navigation was so substantial that we did not need to use subtle quantitative metrics developed for the assessment of human-aware navigation methods [133], [135].

9.2 Comparison of Spatio-Temporal Maps

9.2.1 Two Criteria for Spatio-Temporal Maps Comparison

We claim that human-aware navigation needs to include spatio-temporal maps and use them for the decisions in advance of the navigational task. Therefore, we want to compare the maps in their ability to support navigation systems.

The predictions produced by approaches in our comparison (Section 9.2.3) are real values. They are interpreted as the likelihood of human presence at some position in space and time in spatio-temporal scenarios and as the likelihood of people moving with a specific velocity at some position and time in spatio-temporal-directional scenarios. We created simulated scenarios at different times, where the goal for the robot was to visit 3 distinct places in a university hall. The predictions filled a spatial grid with the predictions, and Dijkstra's algorithm found the cheapest path considering the predictions. We let the system choose the order of the visits in each run which lets the robot follow the predicted pedestrian flow.

Following our previous work [23], we compare the quality of the spatio-temporal map primarily by the expected encounters EE with the associated quantile function $SD(r)$, referred to as service disturbance. The *encounters* are obtained using simulated movement of the robot. The robot follows the path obtained from Dijkstra and continually recalculates people in its predefined vicinity every 10 centimetres. As we hypothesise that the disruption of people is highly correlated with the unexpected replanning of their movement, the contacts with people are weighted w by a mutual robot, and human direction of movement and

referred to as encounters,

$$w = 1 + \frac{v_h \cdot v_r}{||v_h||||v_r||}, \tag{9.1}$$

where v_h and v_r are the velocities of a human and robot, respectively. The weight w follows the idea of the upstream criterion [138] providing us with information about whether the given model correctly predicts the directions of pedestrian flows. The encounters represent the true costs, while the aforementioned method to obtain them represents the true cost function.

During performing the comparison, we encounter the need for the differentiation between models that force the robot to *always* go around the walls and the models that additionally provide information about when it is possible to choose a short path directly through the hall. Therefore, we defined a second true cost function that provided us with the distance travelled by the robot in each run. We denote the secondary measure of the quality as an *expected length* EL and associated quantile function as a *travelled distance* $TD(r)$. It provides information on the ratio of the chosen paths predicted as 'safe' regarding meeting people and the ratio of safe and short paths. Such information can be exploited to estimate the robot's ability to decide on safe and fast traversals of the environment when using a given model.

In general, the patio-temporal maps provide predictions to the navigation systems, while the reactive navigation solves the current situation a robot encounters. As such, the criteria we used in our comparison are not focused on testing specific manoeuvres like in pure human-aware navigation papers, for example [133], [135], but cover the similar principles directly connected to the specialised human-aware manoeuvres. If the scientist focusing on human-aware navigation recognise the need for another, more specialised criterion to compare the predictive ability of maps, they can follow the methodology proposed in Section 8.4.

9.2.2 Dataset Collection

The dataset collection was performed in a hall of approximately $500m^2$ on the UTBM university campus. As shown in Figure 9.4, a Velodyne HDL-32E 3D LiDAR was placed in the reception near the building door to ensure safe 24-hour operation. The spatial placement of the LiDAR was carefully determined

to ensure maximum field-of-view (i.e. approximately $200m^2$) of the hall beyond the glass windows. The raw data of the LiDAR was recorded to ROS *rosbags* 24 hours a day.

Our dataset consists of one month, March 2019, of continuous human detection used as the training set. The testing dataset consisted of 7 days from one week in December. More than 6 million human detections were generated from the LiDAR recordings. They were extracted using the FLOBOT human detection and tracking system [83]. The false positive detections were removed by searching for extraordinarily stable and immovable human detections. Later, we filtered out reflections of people by manual localisation of places producing those reflections. Compared to the dataset used in our previous work [23], the training dataset was twice as long, and the testing dataset did not immediately follow the training period. We also utilised our experience in filtering out the false positive detections, which led to more effective scripts.

Due to the temporal and spatial continuity of data, the dataset reflects people's regular activities, like students entering the building lobby at the beginning of class, leaving after class, and eating and chatting in the lobby during lunch break.

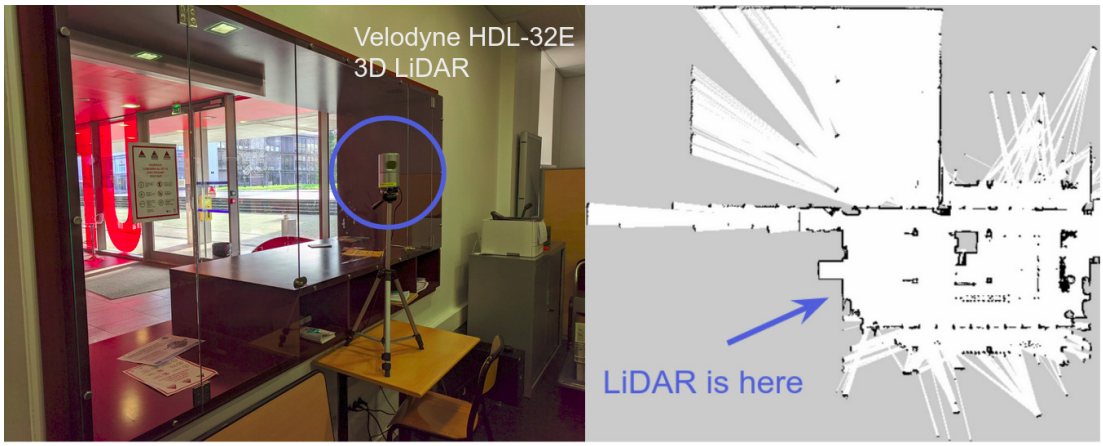

Figure 9.4: Left: A Velodyne HDL-32E 3D LiDAR placed in the reception near the UTBM building door. Right: The occupancy map build by a Toyota HSR robot [137] with a Hokuyo UST-20LX 2D LiDAR.

9.2.3 Approaches in Comparison

Our comparison focuses on the general principles rather than the optimisation of the parameters of the methods tailored directly to the analysed data. The methods found in the literature can be divided into a few overlapping groups of approaches. Compared to the related work section, we excluded a group of methods incorporating short term dynamics and included time series forecasting approach that estimates the number of people over the whole map at a specific time:

1. Spatial-only models, (Section 7.1), that do not take time into account [92], [113],
2. Time series forecasting methods that do not take the structure of the space into account [18],
3. Partially discrete and partially continuous models, (Section 7.3.3), which incorporate continuous models in the cells of predefined grid [57], [63].
4. Methods that model spatial and temporal features separately, (Section 7.3.1), and understand them as independent [40], [119],
5. Continuous spatio-temporal methods, (Section 7.4.2), that model the spatio-temporal phenomena together [16], [20],
6. Continuous spatio-temporal methods, (Section 7.4.1), that model the temporal evolution of continuous spatial model [39], [43].

We implemented various predictive spatio-temporal maps deduced from state-of-the-art principles. The names of the compared approaches are derived from the names of the original methods to make the following text as readable as possible. Our aim was not to systematise the names of the approaches but to maximise the clarity and minimise the time a reader needs to learn and memorise the differences between approaches' internals. These names and abbreviations are written in italics to avoid the reader's confusion. The description of the methods' functioning is simplified, and readers who want to implement them should follow the original proposals. Although some of the compared approaches

were not published, they follow known principles and are not presented here as newly proposed methods.

Spatial Models

As a representative of the discrete spatial models, we choose a spatial grid including the ratio of detected people within each cell to all detections, denoted as *MeanGrid*. We also included *OccupancyGrid* model as a fundamental robotics domain way to represent a map. The occupancy grid captures the environment structure only and neglects dynamic obstacles. Such an approach results in cells with values close to 0 regardless of the number of people moving through them. The continuous spatial model is represented by an expectation-maximisation algorithm for fitting a mixture of Gaussian models [139] with its probabilistic prediction, referred to as *GMM*.

Time Series Forecasting

We included 3 time series forecasting methods recently applied to forecasting human presence and pedestrian flows. Those methods can be viewed as spatio-temporal models, whose spatial map is *OccupancyGrid*. Frequency Map Enhancement *FreMEn* [5] is derived from the Fourier transform and as it was initially proposed for binary data containing binary states. It had to be reimplemented for data that do not contain states but events. It should be noted that although the computational effectiveness of our implementation was greatly enhanced, the predictive ability was lowered due to the inaccuracy of the estimation of heights of amplitudes. Authors of [16] proposed a method for time series forecasting referred to as hypertime. Based on the definition, we incorporated its two variants, *HyT* and *WHyTe*. Those methods exploit the ability of *FreMEn* to detect prominent periods in data. They project time into the multidimensional vector space forming a Clifford torus, and apply *GMM* with one component to the projected data. The most prominent periods are gathered iteratively by applying *FreMEn* to the series of errors between reconstruction and the measurements. They differ in the metric used for the distance calculation. While *Hyt* uses Euclidean metric, *WHyTe* uses a metric derived from cosine metric, Section 9.3.1. We also wanted

to include Prophet [85] as a representative of a time series forecasting method that cannot only analyse time series with an unequal step between measurements but can also model trends. However, we faced difficulties applying Prophet on such extensive data with predictions too far into the future, and we subsequently removed it from the list of the compared methods. The only discrete forecasting method used in the comparison was a histogram over the week-lengthy time window with 168 bins, each covering an hour, *HistWeek*.

Continuous Models in the Grid Cells

Every aforementioned forecasting method can be employed in the spatial grid forming spatio-temporal models similar to the original idea of the 'map enhancement' in [55]. We included into the comparison *FreMEnGrid* as a representative of continuous ones in temporal domain and *HistWeekGrid* as a fully discrete spatio-temporal model. We omitted *HyTGrid* and *WHyTeGrid* as their computational demand in the stage of prediction was considerably higher than of the previous two, and the straightforward optimisation of code would violate the design of the testing environment. The third not fully continuous model included is *time_window_GMM* [119] which expands *HistWeek* with *GMM* connected to each cell. Its prediction then consists of *GMM* predictions in each cell weighted by *HistWeek* forecast.

Independent Modelling of Spatial and Temporal Features

Modelling space independently to time can sound illogical. On the other hand, considering robotic topological maps, we can assume that different nodes will show standardised behaviour that is sometimes more relevant and sometimes almost unnoticeable. We included 7 combinations of spatial and temporal models, whose predictions are multiplication of their independently modelled components: *HistWeek_X_GMM* with a discrete temporal and continuous spatial model, three models including continuous model of time and discrete model of space, *FreMEn_X_MeanGrid*, *HyT_X_MeanGrid*, and *WHyTe_X_MeanGrid*, and three continuous maps *FreMEn_X_GMM*, *HyT_X_GMM*, and *WHyTe_X_GMM*.

Continuous Spatio-Temporal Models

The continuous maps that model the time and space together show significant differences in the computational demand. The most demanding are spatio-temporal models based on hypertime. Although they proved their ability to model space-time effectively [20], [21], [23] by the mean of the model quality and size of the model, their use on real robots is questionable since their iterative nature consumes many recourses. We included *FreMEn_HyTS_clusters* derived from the *HyT* algorithm and *FreMEn_WHyTeS_clusters* derived from the *WHyTe* algorithm. Contrary to their time series forecasting variants *HyT* and *WHyTe* that iteratively choose the best periodicity, our implementation of those spatio-temporal variants first applies *FreMEn* to the data to gather the most prominent periods and then project data to the hypertime-space. Both apply *GMM* over the projected data to get the model regarding respective metrics. We also included 'lightweight' versions that apply *GMM* using only one component and we denoted them as *FreMEn_HyTS* and *FreMEn_WHyTeS* respectively.

Temporal Evolution of Continuous Spatial Model

We also included continuous spatio-temporal approaches that fit the space's structure and then model the temporal changes, similar to [43]. *WHyTened_kMeans* divides the space into disjunct subsets using the k-means clustering algorithm and applies *FreMEn_WHyTeS* to every part. The second method, *HyTted_GMM*, applies *GMM* to the space and its every component defines the time series by the measurements matched with it. Those time series are then modelled with *HyT*. The predictions in both methods are calculated as a multiplication of the clustering algorithm prediction and the spatio-temporal or temporal model prediction. The *WHyTened_kMeans* can be understood as multiple disjunct spatio-temporal models, while *HyTted_GMM* represent continuous spatio-temporal models where each component has its specific temporal characteristics.

9.2.4 Testing Environment Setup

The experiments were conducted in a general testing framework with a unified interface available for individual methods. The framework consisted of four stages: training, prediction, pathfinder, and simulation.

The predictors were trained over spatio-temporal and spatio-temporal-directional data during the training stage. There are approximately six million detections of people in the training data. We also included scenarios where the training dataset includes only one per mile randomly chosen detections to provide insight into how robust different methods are to data sparsity.

During the prediction stage, the framework created a spatio-temporal grid over the testing week. The week was broken into 40 second windows. For each of these windows, the area of approximately 200m^2 was broken into a grid of 0.25m^2 cells. Each spatial cell consisted of 8 directional cells. The predictors filled the directional cells with their estimations of how likely a human would go at that time and place in that direction.

The pathfinder applies the Dijkstra search algorithm for each time window, where the cost of a transition is determined by prediction from the tested method. The cost for Dijkstra at a specific position is calculated as an estimation of the cost of an encounter multiplied by the distance to the next state. The paths chosen in each time window are saved together with their lengths and costs. Note that the overall length of paths planned during the whole testing week was hundreds of kilometres.

The last stage of the evaluation is a self-written kinematic simulator. The simulation is done per time window for each model while saving the intermediate results for correctness verification and recording the runtime duration. The robot moves by the chosen path at a predefined speed. The simulator calculates weighted encounters with human detections on the robot's trajectory. The sum of weighted encounters is saved together with length and path cost, which allows for criteria application.

The framework allows a parametrization of the robot's movement and goal placement for planning. The experiments were evaluated with the following settings: the robot radius is 1 meter, which covers the hypothetical robot radius

together with a human radius, and the robot speed is $0.5ms^{-1}$.

Some compared approaches also required parameter setting. The adjustment of the parameters was not part of our investigation. The number of components for algorithms applying *GMM* (and k-means) was set to 10 except for *FreMEn_HyTS_clusters* and *FreMEn_WHyTeS_clusters* where we set 5 components (to shortened the computational time). The number of periodical components was set to 5 in every method employing *FreMEn*. We used a computational (floating point) precision of 64 bits to compare and order continuous values.

9.2.5 Results of Comparison

The comparison of methods applied to spatio-temporal data can be seen in the subgraphs in Figures 9.5 and 9.5. The predictors predicted how likely the robot was to meet a person at a given space and time coordinates. In each row, we have two subfigures, with the left one depicting a graph of the dependence of the service disturbance on the servicing ratio and the right one the length of the path on the servicing ratio.

In the top row of Figure 9.5, we compare the quality of the predictions of the spatial-only models. None of them considers time, which leads to a random ordering of the costs of the paths. Therefore, the dependence of the service disturbance on the servicing ratio grows linearly. Similarly, as the space models do not change in time, the travelled distances grow also linearly. The *OccupancyGrid* predicts only zeros, and therefore its expected length of paths is the shortest possible out of all models. *GMM* forces the robot to go closer to the walls than *MeanGrid*, which leads to longer paths and lower expected encounters.

In the second-from-top row, we compare models that take into account only time, which means that those models can order the costs of the paths, but they always use the shortest path possible. Such predictions lead to linear growth of the travelled distance identical to the *OccupancyGrid*, but non-linear growth of the service disturbance. All methods can estimate the safest 10% of the time windows to provide the service, while the most successful, *HyT*, can predict 50% of the safest time windows. *WHyTe* and *FreMEn* perform similarly and *HistWeek* is closer in performance to the *HyT* than to the other two models. The rank of

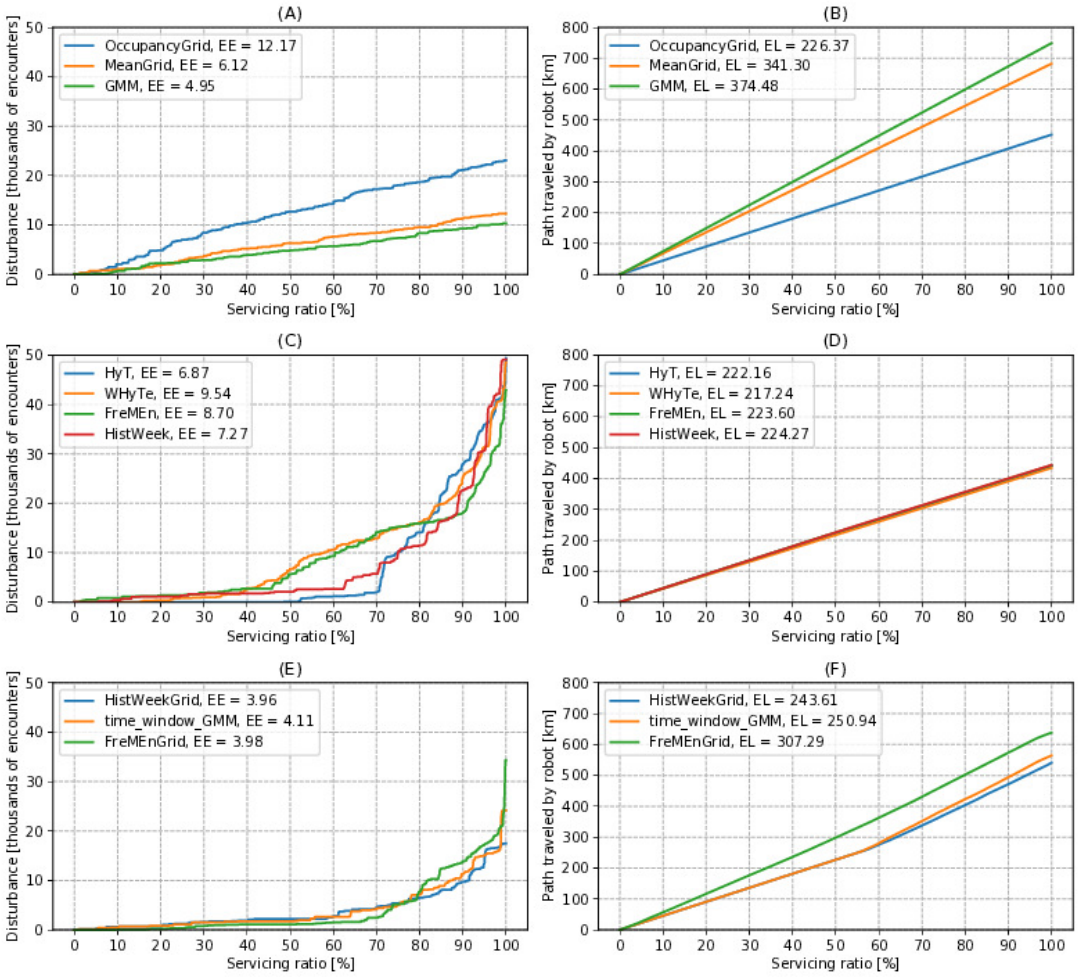

Figure 9.5: Comparison of different models in the spatio-temporal scenario. The left column of graphs depict service disturbances, the quantile functions of the ordered human-robot encounters. The right column of graphs depict traveled distances, the quantile functions of lengths of paths. The labels in graphs consist of names of models and the expected values.

the quality of prediction of the compared methods changes at 70% of servicing ratio. At a servicing ratio of over 90%, *FreMEn* performs best.

In the third row, we compare spatio-temporal models that are fully or partially discrete. *HistWeekGrid* performs similar to *time_window_GMM*. Both models force the robot to follow the shortest paths until almost a 60% servicing ratio. Then, their predictions start to force robots to follow a longer path, evading the

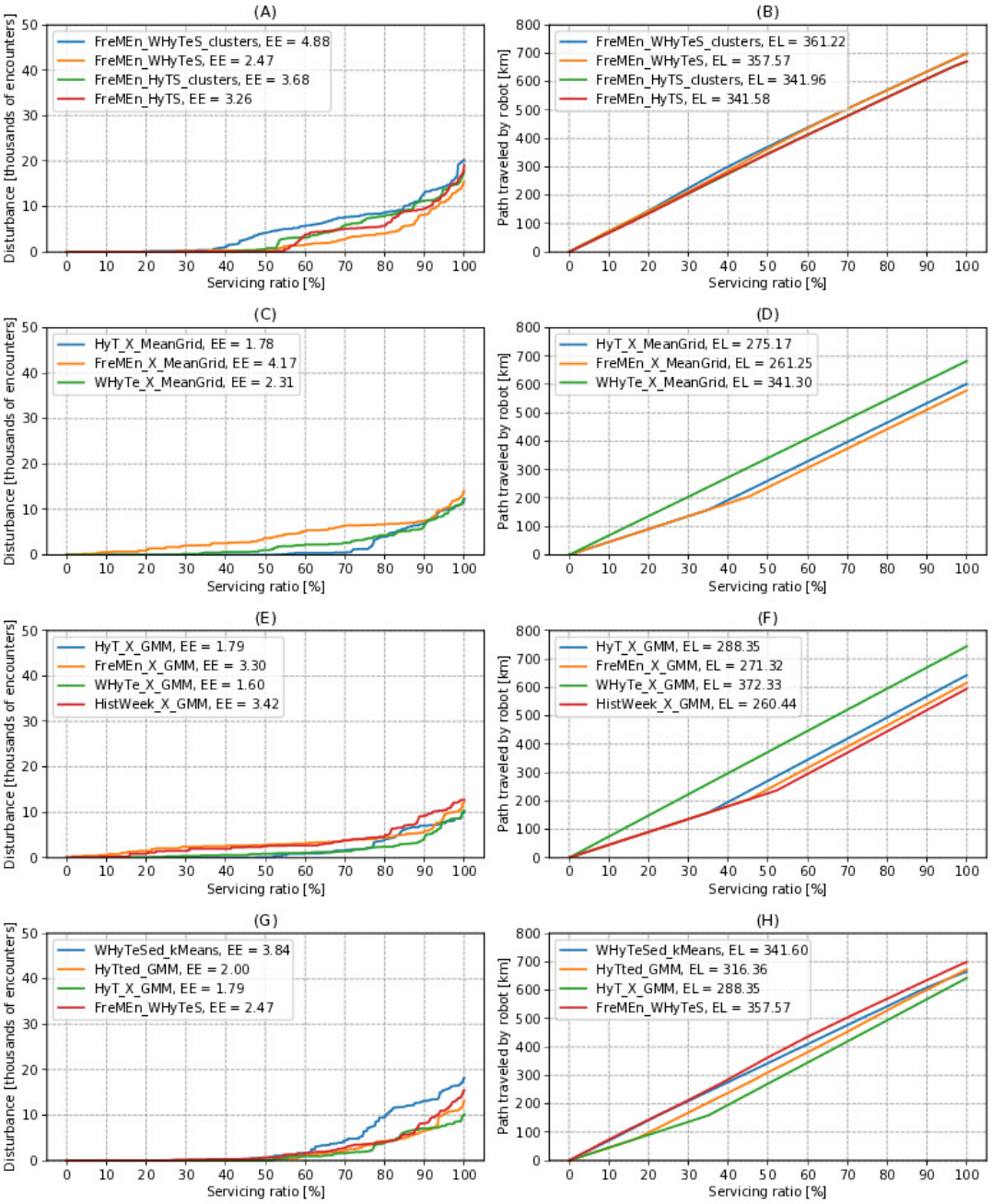

Figure 9.6: Comparison of different models in the spatio-temporal scenario - continuation of Figure 9.5.

most critical parts of the hall. Similar to the spatio-only models *GMM* and *MeanGrid*, *time_window_GMM* forces robot to travel longer paths than *HistWeekGrid*.

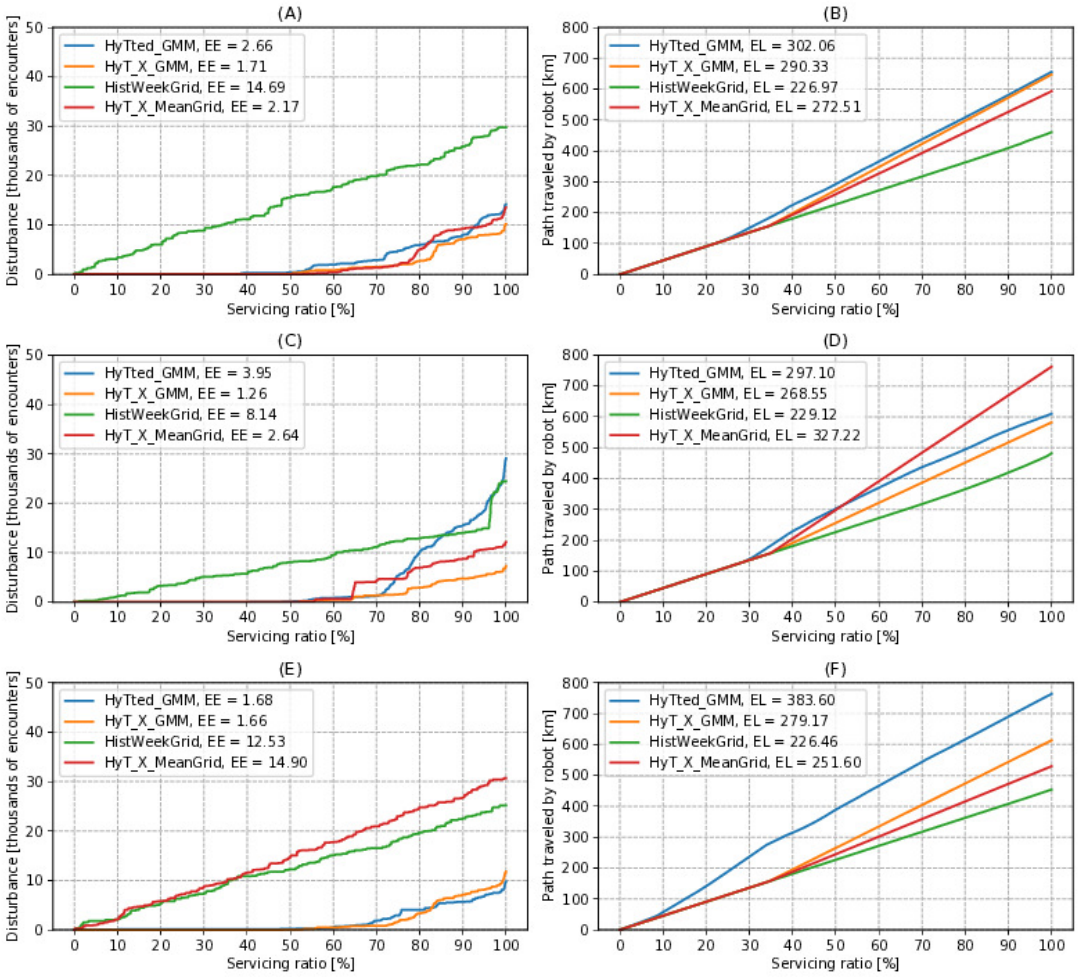

Figure 9.7: Comparison of selected models under different scenarios. The left column of graphs depict service disturbances, the quantile functions of the ordered human-robot encounters. The right column of graphs depict traveled distances, the quantile functions of lengths of paths. The labels in graphs consist of names of models and the expected values. First row: models trained over the sparse, spatio-temporal data. Second row: models trained over the spatio-temporal-directional data. Third row: models trained over the sparse, spatio-temporal-directional data.

The *FreMEnGrid* model is better than the other two models in scheduling until more than 70% of servicing ratio, but at higher servicing ratios, it loses its edge.

In Figure 9.6, the top row provides a comparison of the continuous spatio-temporal models. All of them have a linear growth of the travelled distance, and

neither can provide predictions that would let the robot travel by the shortest path. Moreover, the spatio-temporal model *FreMEn_WHyTeS_clusters* modelling five Gaussians performs in expected encounters worse than the similar model *FreMEn_WHyTeS* with only one component. The loss of quality between *FreMEn_HyTS_clusters* and *FreMEn_HyTS* is not so drastic. However, the evaluation of the continuous spatio-temporal models that apply *GMM* over the hypertime provides us with information that the robot always follows a long path. On the other hand, *FreMEn_WHyTeS* provides the best expected encounters between methods already compared.

The following two rows of graphs compare the spatio-temporal methods that model space and time separately as independent phenomena. Many of those methods perform surprisingly well, especially in choosing the right time to go by the shortest path. *HyT_X_MeanGrid* and *HyT_X_GMM* can lead the robot by the shortest path in 35% of time while evading almost every encounter in 50% of the time. Other methods exchange the performance in the length of the path for safety. *FreMEn_X_MeanGrid*, *FreMEn_X_GMM*, and *HistWeek_X_GMM* were unable to evade encounters.

Based on the results above, we combine the good qualities of the previously tested ones. As the *HyT_X_MeanGrid* and *HyT_X_GMM* perform very well in raw combinations of spatial and temporal models and extending *FreMEn_HyTS* with multiple components *FreMEn_HyTS_clusters* did not result in a positive change of the quality, we combined *HyT* and *GMM* in a way that every component has its own temporal model. The method is denoted as *HyTted_GMM*. A similar process led us to create *WHyTened_kMeans*. *FreMEn_WHyTeS* performed quite well in the sense of expected encounters but did not find the shortest path. This behaviour led to the idea of "natural" divisions of space by k-means clustering while applying the successful spatio-temporal method to the subsets. The comparison of these two models with *HyT_X_GMM* and *FreMEn_WHyTeS* can be seen in graphs in the bottom row of Figure 9.6. We can see that the disturbance distribution functions of *HyT_X_GMM*, *FreMEn_WHyTeS*, and *HyTted_GMM* are very similar. Out of these three models, *HyT_X_GMM* provides the best expected encounters, and *FreMEn_WHyTeS* the worst. *WHyTened_kMeans* performs much worse than others, and it did not even get the ability to provide the robot predic-

tions that let it go through the environment with the shortest path. *HyTted_GMM* acquired the ability to predict short and safe paths more than 15% of times while being a fully continuous spatio-temporal method.

We choose the three best-performing methods for the comparison in the subsequent scenarios, *HyT_X_GMM*, *HyT_X_MeanGrid*, and *HyTted_GMM*, together with *HistWeekGrid* which provided the best ability to lead the robot by the shortest path while performing exceptionally in evading encounters. On the top row of Figure 9.7, we provided a comparison of chosen methods in the spatio-temporal scenario when we randomly selected only one per mile of human detections from the training dataset to train the models. The *HistWeekGrid* lost its ability to provide predictions. In the previous scenario, predictions in each cell were calculated from approximately 2000 values, while in this scenario, only from 2. The other three methods performed well. The *HyTted_GMM* is the worst of the rest of the methods in the range of $50 - 85\%$, but it can choose 25% of the safe and short paths, which is almost similar to the other two functional models. Moreover, *HyTted_GMM* and *HyT_X_MeanGrid* found shorter "safe" paths than *HyTted_GMM*, which can be deduced from different inclinations of the traveled distances.

In the second row of graphs in Figure 9.7, we compare the same four methods in a spatio-temporal-directional scenario with the full amount of detections. The methods predict when to go, what path to choose, and what direction is the best to go with the flow of people. The *HistWeekGrid* provides us with an inferior model that can model only the most frequent parts of the spatio-temporal-directional grid. The *HyT_X_GMM* and *HyTted_GMM* perform similarly until up to a 70% servicing ratio, where *HyTted_GMM* choses a more risky and shorter path to follow. After following the shortest path, the *HyT_X_MeanGrid* chooses the longest path out of the three well-performing models, which means it led the robot in close vicinity to walls. The best performance by the mean of evading people is provided by *HyT_X_GMM*.

In the bottom row, we can see the performance of the methods when providing predictions in the spatio-temporal-directional scenario obtaining only one per mile of samples for the training. The grid-based methods *HistWeekGrid* and *HyT_X_MeanGrid* were not able to provide a prediction. The continuous mod-

Table 9.2: Comparison of the approaches using popular metrics

	χ^2d	AUC	RMSE	MLL	PCC
Section 9.2.3					
MeanGrid	1.319811e+06	0.678204	1.131719e-07	-5.964167e-07	0.004277
OccupancyGrid	**1.077451e+05**	0.500000	**1.107762e-07**	**-5.714333e-07**	0.000000
GMM	1.895740e+06	0.661219	1.119760e-07	-5.838786e-07	0.006606
Section 9.2.3					
FreMEn	1.088325e+06	0.708411	1.112586e-07	-5.764210e-07	0.017761
HyT	1.157291e+06	0.786832	1.115330e-07	-5.792680e-07	0.018118
WHyTe	1.896955e+06	0.735896	1.128428e-07	-5.929532e-07	0.021518
HistWeek	9.583215e+05	0.770708	1.112834e-07	-5.766781e-07	-0.010873
Section 9.2.3					
FreMEnGrid	7.824628e+05	0.724162	1.113846e-07	-5.777281e-07	0.019364
HistWeekGrid	4.436296e+05	0.790896	1.109813e-07	-5.735518e-07	-0.019459
time_window_GMM	8.572676e+05	0.811935	1.256092e-07	-7.347102e-07	0.014281
Section 9.2.3					
HistWeek_X_GMM	9.599045e+05	0.817996	1.112969e-07	-5.768189e-07	-0.011000
FreMEn_X_MeanGrid	7.703107e+05	0.745279	1.112063e-07	-5.758796e-07	0.006380
HyT_X_MeanGrid	8.180353e+05	0.816573	1.112772e-07	-5.766142e-07	0.005943
WHyTe_X_MeanGrid	1.318568e+06	0.768864	1.117257e-07	-5.812720e-07	0.009677
FreMEn_X_GMM	1.088325e+06	0.739453	1.111231e-07	-5.750178e-07	0.015433
HyT_X_GMM	1.154622e+06	**0.829502**	1.112700e-07	-5.765397e-07	0.013693
WHyTe_X_GMM	1.895740e+06	0.774540	1.141287e-07	-6.065450e-07	0.024228
Section 9.2.3					
FreMEn_HyTS_clusters	1.578800e+06	0.808032	1.121873e-07	-5.860852e-07	0.023457
FreMEn_WHyTeS_clusters	1.857979e+06	0.788587	1.134001e-07	-5.988245e-07	0.021970
FreMEn_HyTS	1.896951e+06	0.814097	1.140148e-07	-6.053347e-07	**0.032425**
FreMEn_WHyTeS	1.896950e+06	0.772722	1.135045e-07	-5.999280e-07	0.027456
Section 9.2.3					
HyTted_GMM	1.152000e+06	0.817629	1.113172e-07	-5.770285e-07	0.013649
WHyTened_kMeans	1.870547e+06	0.783438	1.136826e-07	-6.018122e-07	0.026585

els perform similarly in expected encounters but differ mainly in the paths they choose. The *HyT_X_GMM* retain its behaviour from the previous scenario. However, *HyTted_GMM* started with the shortest path in the first 10% of planned services; then, it found out encounters can emerge and reacted to that 'hysterically' forcing the robot to go around the walls. At about 30% of the servicing ratio, it started to follow similar paths to *HyT_X_GMM*.

9.2.6 Comparison of Approaches Using Popular Methods

For the sake of comprehensiveness, we compare the approaches to spatio-temporal mapping using measures of quality that appeared in the referenced literature (Section 7.5). The most favoured measures of the quality of maps are mean square error, rooted mean square error (RMSE), mean negative log-likelihood, and mean

Table 9.3: Correlation of the popular metrics' results

	χ^2d	AUC	RMSE	MLL	PCC
χ^2d	1.00	0.29	0.14	-0.13	0.68
AUC	0.29	1.00	0.22	-0.21	0.18
RMSE	0.14	0.22	1.00	**-1.00**	0.26
MLL	-0.13	-0.21	**-1.00**	1.00	-0.25
PCC	0.68	0.18	0.26	-0.25	1.00

log-likelihood (MLL). We choose RMSE and MLL as their representatives in the comparison. We also included the area under a receiver operating characteristic curve (AUC) [140] as a complex representative of popular measures derived from the confusion matrix. We also included Chi-Square Distance (χ^2d) [141], which appeared in articles in the last few years as an alternative to RMSE. Some measures of the quality of maps appeared in the literature very rarely. Many of them are hard to implement, and few are poorly defined. We included Pearson correlation coefficient (PCC) [142] as a representative of rare measurements because we found it the most traditional and feasible to implement.

Table 9.2 provides the results of χ^2d, AUC, RMSE, MLL, and PCC. The slight differences between the values representing the quality of the models led us to compare them with very high precision representations, even for our data consisting of millions of detections. The best spatio-temporal map is OccupancyGrid according to χ^2d, RMSE, and MLL. We interpret the result so that the detection of humans is quite a rare event. Therefore the best predictor is the one that predicts almost zero probability everywhere. Other maps also consist of small valued predictions but higher than 'almost zero'. As the detection is one point in space and time, the (hypothetically correct) predictions in the vicinity of the detections raise the error of the whole model. Similarly, the output of PCC for all the approaches gave us almost zero correlation between maps and the detections. It is not reasonable to deduce any conclusions from such small values. Contrary to previous ones, AUC provided a very reasonable ranking. We can deduce which of the approaches gave us good predictions. However, the interpretability of the outputs is very limited.

We also included the correlation between the metrics' results, Table9.3. RMSE and MLL are highly correlated, providing almost interchangeable information.

There is also a rather significant correlation between $\chi^2 d$ and PCC. The correlation between AUC and other techniques is insignificant.

9.3 Comparing State-of-the-Art Methods

The proposed methodology allowed us for State-of-the-Art methods comparison. We contacted scientific teams focusing on spatio-temporal or directional mapping. As the goal was not to portray one of the algorithms as the best but to discover the good and bad attributes of different approaches, the teams were debugging codes and tuning the models during a few-days-long preparatory stage. Every team received the training data and was provided with the current results on a rolling base. In the end, every team provided a model of its choice and described it.

9.3.1 Methods Involved in the Experiments

WHyTeS

The abbreviation WHyTeS comes from *Warped Hypertime Space*. Its main idea is the projection of time onto a Clifford torus. From a mathematical point of view, the difference between HyT (Hypertime) and WHyTe (Warped Hypertime) lies in a different calculation of distances between projected vectors. While HyT calculates Euclidean distance in Hypertime, WHyTe calculates Euclidean distance on a surface of a Clifford torus constructed accordingly to Hypertime. WHyTeS, in the default setting, uses only $c = 1$ cluster. The method combines density estimation on a Clifford torus and the spectral decomposition derived from FreMEn [5].

WHyTeS models the flows of people in space-time $(t, x, y, \dot{x}, \dot{y})$, but for the testing method, we need to estimate the weights in the space of (t, x, y, ϕ). Therefore, we create a histogram of predictions over $(t, x, y, \dot{x}, \dot{y})$ and calculate a sum of predictions in cells that lie between $\phi - \frac{\pi}{8}$ and $\phi + \frac{\pi}{8}$ in every (t, x, y).

CLiFF-Map

Circular Linear Flow Field Map (CLiFF-Map) [73] is a technique for encoding motion patterns probabilistically. The probability density function (PDF) representing the CLiFF-Map is a semi-wrapped Gaussian mixture model (SWGMM) for each point in the map.

The edge weights for the Dijkstra graph are computed using the Extended Upstream Criterion (EUC) proposed in [74]. The EUC penalises paths that do not conform to the underlying CLiFF-Map distribution. In the Dijkstra graph, the weight of each directed edge is computed using the CLiFF-Map components at the node's location to which the edge leads.

STeF-Map

Spatio-Temporal Flow Map (STeF-Map) [64] is a representation that models the likelihood of motion directions on a grid-based map by a set of harmonic functions, which capture long-term changes in crowd movements over time. The underlying geometric space is represented by a grid, where each cell contains K temporal models, corresponding to K discretised orientations of people-motion through the given cell over time. The temporal models, which can capture patterns of people's movement, are based on the FreMEn framework [5]. The edge weights for the Dijkstra graph are calculated using the EUC analogous to the CLiFF-Map application.

Other Models

In addition to the above-listed state-of-the-art methods, we include for comparison three historical average-based models and one time series forecasting model, all of which are commonly used as baseline methods. The average-based models used were: Mean, which predicts an average of its past measurements for each spatial segment, and two histograms, Hist24 and Hist168, that describe the average day and week. Both histograms split their period into one-hour-long segments and compute the average value for each segment. An industrial time series forecasting tool Prophet [85] was trained with measurements condensed into one-hour-long time steps. All these methods were trained on individual spatial

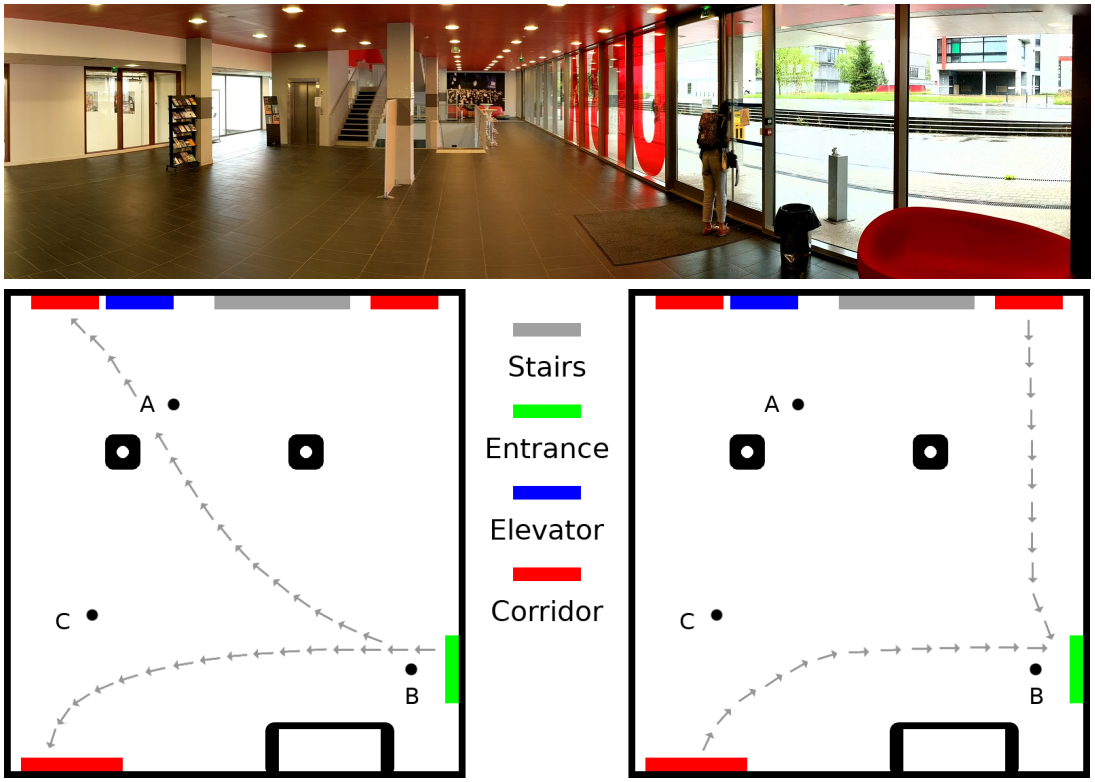

Figure 9.8: Location where the experiments were performed. Photo and most prominent pedestrian flows during the morning (left) and evening (right).

segments without accounting for spatial relations. The last model is denoted as *Occupancy grid*. As this map does not include 'walls', which are static and define borders of an area we are focusing on, this purely static model "predicts" at all cells value 0, which is equivalent to "free", and allows for the planning without forecasting the flow.

9.3.2 Evaluation dataset

The approaches described above were evaluated using a dataset collected in Building M at the University of Technology of Belfort-Montbeliard (UTBM). The data recording was performed by a Velodyne HDL-32E 3D lidar, using a reliable person detection method [83]. During the data collection, the lidar remained stationary in the reception room near the entrance of the building, which allowed to scan the main activity area of the hall, covering a total area of around 500 m^2 (Fig-

ure 9.8). The data collection was performed on a full 24/7 basis for three months in 2019, and the dataset also contains semantic information, including positions of the entrance, elevator, stairs, and corridors.

We restricted the area of detections to cover only the main entrance hall (approx. 150 m^2). The training dataset includes eleven working days from March 2019, when the students regularly attended classes, and the test dataset is the first Monday in April 2019. Each day contains approximately 300 000 human detections. Every detection is represented by a vector $(t, x, y, \dot{x}, \dot{y}, \phi, v)$ – time of detection, $2d$ position, $2d$ velocity, orientation, and speed of the detected human. Note that $(\dot{x}, \dot{y})$ and (ϕ, v) are mutually convertible.

The 3D lidar was mounted in a reception office at a height of about 1.2 m, providing a good overview of the environment for data collection (Fig. 9.8). The 3D point cloud generated by the Velodyne lidar was divided into separate sets using an adaptive clustering method [83]. These clusters were then classified as human or non-human using a support vector machine (SVM), and 2D positions of people were processed by a multi-target tracking method [89] based on a combination of Unscented Kalman Filter (UKF) and Nearest Neighbour Joint Probability Data Association method (NNJPDA). The trajectories calculated by the tracker were then examined manually, and outliers (e.g. static objects classified as people) were removed from the dataset.

9.3.3 Details on path planning

We provide the robot with a set of navigation scenarios starting at t_i. We do not require that the robot performs every navigation task, but only a certain fraction of them. With a lower servicing ratio, the robot has more freedom to decide when to navigate through the area and when not. This reflects the situation when the robot has to perform a certain number of tasks during the day, but it can choose the best times to do so [60], [143].

The scenario is inspired by a security robot procedure, where the robot has to visit a few predefined locations. In our case, there are one starting and finishing position (A) and two goal positions (B, C); see Figure 9.8. The robot decides the order in which it will check the positions B and C, i.e. if it visits the locations in

order (A, B, C, A) or (A, C, B, A). This decision is deduced from the mean of the costs of each possible path – the robot chooses the path from the prediction of pedestrian flows.

To include people's personal space and the robot's size in results, the path planning method assumes that the robot's social distance radius is 1 m, and encounter detection is triggered every 0.1 m along its trajectory. The robot speed was set to 0.5ms^{-1}. The robot performed 597 patrols during one day between 6:30 am and 8 pm starting every 80 s.

9.3.4 Chosen Criteria

Total Encounters and Expected Encounters

Except for the expected encounters criterion EE, we count all blind robot-human encounters, referred to as *total encounters* (TE) during all planned passages through the environment,

$$TE = \sum_{i=1}^{p} e_i. \tag{9.2}$$

The value TE reflects the model's overall ability to support unobstructive planning by preventing the disruption of pedestrian flows. Lower total encounters mean a better human flow model of the environment.

The main reason for introducing both criteria, TE and EE, lies in the difference between time-averaged and time-sensitive models. For time-averaged models, EE has no meaning, and its value lies around $\frac{1}{2}$TE (biased by the random ordering of services). Although there is apparent reason to compare time-sensitive models by TE, their sensitivity to the time-dependent changes should be compared by their ability to correctly predict the (relative) number of people at specific times, which leads to EE.

Other Criteria

In addition to the proposed criteria to measure the quality of models, we included RMSE and Chi-square between the cost map predicted by the model and the ground truth obtained from the testing dataset, following the earlier comparison

of human flow models [21]. RMSE is widely used in time series forecasting:

$$RMSE = \sqrt{\frac{1}{p \cdot n \cdot a} \sum_{i=1}^{p} \sum_{q=1}^{n} \sum_{b=1}^{a} (x'_{i,q,b} - y'_{i,q,b})^2}, \tag{9.3}$$

normalised as a distribution over the testing space-time

$$\sum_{i=1}^{p} \sum_{q=1}^{n} \sum_{b=1}^{a} x'_{i,q,b} = \sum_{i=1}^{p} \sum_{q=1}^{n} \sum_{b=1}^{a} y'_{i,q,b} = 1, \tag{9.4}$$

and the Chi-square distance is used to compare histograms:

$$\chi^2 distance = \sum_{i=1}^{p} \sum_{q=1}^{n} \sum_{b=1}^{a} \frac{(x''_{i,q,b} - y''_{i,q,b})^2}{(x''_{i,q,b} + y''_{i,q,b})}, \tag{9.5}$$

normalised over the angles in every cell

$$\sum_{b=1}^{a} x''_{i,q,b} = \sum_{b=1}^{a} y''_{i,q,b} = 1 \tag{9.6}$$

where p is the number of scenarios, n is the number of positions, a is the number of angular bins, $x'_{i,q,b}$, resp. $x''_{i,q,b}$ is the normalised value of estimated cost for angle b at position q in scenario i, and $y'_{i,q,b}$, resp. $y''_{i,q,b}$ is the normalised value obtained from the ground truth at the identical position.

9.3.5 Hypothesis on Results of Models

We propose two criteria to quantify the quality of models, total encounters TE and expected encounters EE. The natural assumption is that time-averaged models (CLiFF-Map, Mean) should have higher values of TE and EE compared to time-sensitive ones (WHyTeS, STeF-Map, Prophet, Hist24, Hist168), and the difference should be much more significant when comparing EE. Models using continuous time-forecasting (WHyTeS, STeF-Map, Prophet) should have a lower EE than discrete ones (Hist24, Hist168). A simple Occupancy grid (planning without forecasting the flow) should achieve the significantly worst rating, while specialised CLiFF-Map should indicate better results than Mean. RMSE and

Table 9.4: Performance of the Models

Evaluated method	Total encounters	Expected encounters	χ^2 distance	Model error (RMSE)
Occupancy grid	7123	3993	8866	3.65e-06
Mean	4185	2353	147769	4.08e-06
CLiFF-Map	3105	1480	144486	3.76e-06
Hist168	5415	2857	80846	3.65e-06
Prophet	2637	596	59405	3.66e-06
WHyTeS	2835	385	82384	3.67e-06
STeF-Map	1548	332	70535	3.66e-06
Hist24	2898	323	49981	4.02e-06

Chi-square should not reflect the known attributes of the compared models and should provide misleading results.

9.3.6 Evaluation results

The evaluation results are summarised in Table 9.4. TE is supposed to measure the similarity between a given model and the spatial dynamics of the environment, while EE measures the similarity between a model and the spatio-temporal dynamics of the environment. We can see that TE of Mean is significantly lower than TE of Occupancy grid. The specialised CLiFF-Map is the best out of the three models, which do not model the time. EE of these methods is close to one-half of the total encounters, as expected. CLiFF-Map is better than Mean in all the criteria presented in Table 9.4.

Except for Hist168, all models that predict the spatio-temporal dynamics of the environment reach significantly lower values of EE than the time-averaged models. Hist24 achieved the best EE, while Hist168 failed to learn a usable model. To understand this result, we have to point out that the testing data came from a working day and Hist24 benefits from that in the comparison. We can speculate with confidence that its result during the weekend would be weak - there are no lectures, and human flows indicate very different behaviour. On the other hand, Hist168 can incorporate the difference between working days and weekends. However, a large number of temporal bins, along with the short training period, led to a poor model. WHyTeS and STeF-Map covered three

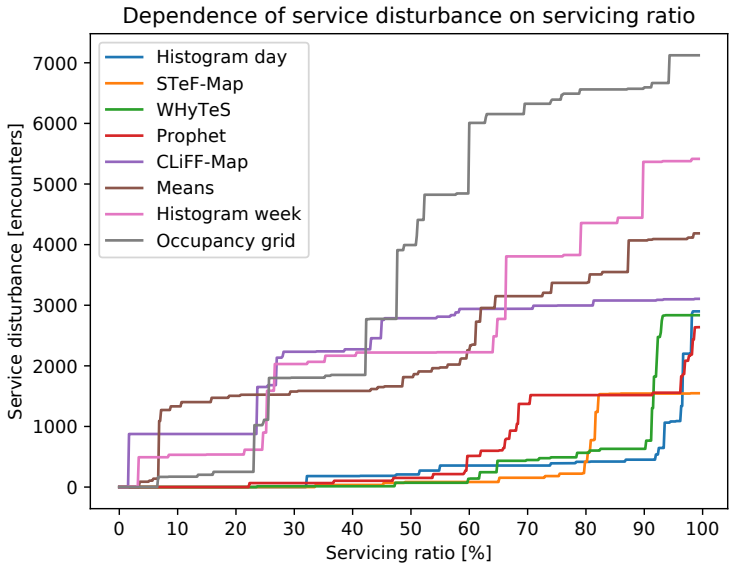

Figure 9.9: The dependence of the number of encounters (service disturbance) on the frequency of the traversals (servicing ratio) achieved by different pedestrian flow models.

dominant periods in the data, including expected day and week. WHyTeS also covered a period corresponding to the length of lectures during the working days; see Figure 9.3. Prophet showed that, for relatively short-term predictions, we can use regular time series forecasting methods and expect good results.

It is much harder, if not impossible, to interpret the characteristics of the predictors in the above way when analysing the results by RMSE or the χ^2 distance. Moreover, the traditional comparison between Occupancy grid and the other two time-averaged models is counter-intuitive and misleading. This comparison supports our hypothesis that the proposed criteria characterise the predictive power of the spatio-temporal human flow models in a useful way. The popular and most common tools do not necessarily indicate how useful the analysed models are for unobstructive and safe path planning.

More in-depth insight into the dependence of the service disturbance on the servicing ratio (i.e. the dependence of encounters on the frequency of the traversals) is provided in Figure 9.9. The graphs indicate that the service disturbance achieved by the time-averaged models, which neglect the temporal variations of

the flows, scales linearly with the servicing ratio. However, time-aware pedestrian flow models can identify times when the given area is crowded, and traversing through it will result in many encounters. Thus, these methods can choose not to enter the areas during these busy times if allowed.

In Figure 9.9, the graphs indicate decreases of the service disturbance by about 80% for time-aware methods allowing the robot to drop 10-20% of tasks. Compared to traditional models, which capture the environment structure only, the well-trained flow representation can reduce the number of encounters by one order of magnitude at servicing ratio 90% and by an order of two magnitudes at servicing ratio 50%.

The results also indicate that considering the spatio-temporal layout of the pedestrian flows has a tremendous impact on the number of people encountered, see TE of STeF-Map in Table 9.4, when compared to the human flow uninformed baseline (Occupancy grid), which characterises only the static structure of the environment and associates each free cell with a fixed cost. Consider that traditional criteria indicate the opposite.

10 Summary

Derived from our experiences with long-term deployments of autonomous robots, we hypothesised that respecting essential human habits improved robot efficiency and acceptability in the long term. Our field robotic experiment in the natural, human-populated environment supported that hypothesis. We showed that our mobile robot, which took into account pedestrian flows and their changes over time, was considered less disturbing. It strongly indicated that scheduling and planning the tasks using sophisticated temporal modelling methods like WHyTeS and STeF-Map lead to more socially acceptable behaviour of a service robot. We also showed that human flows could be forecasted even months after they were learned due to their periodic nature.

We studied spatio-temporal modelling methods and provided their comprehensive overview. As those methods came from different fields and were intended for different usages, it was impossible to compare their performance. However, we proposed a methodology for defining a miscellaneous collection of criteria for evaluating diverse spatio-temporal maps. As an example related to our expertise, we compared simplified approaches derived from the studied models in their ability to provide beneficial forecasts for long-term deployments of human-aware autonomous robots.

We categorised known approaches to spatio-temporal mapping, defined criteria for their comparison following the proposed methodology, and compared them in a simulation built on real-world data. The results showed that simple time series forecasting methods could not support path planning but could be used to avoid peak times. The spatial-only maps allow the construction of paths that avoid potentially crowded areas, but these paths are unnecessarily long when people are unlikely to be present. Discrete spatio-temporal maps can provide rea-

sonable predictions but require a relatively large amount of data. They are also memory inefficient and not scalable enough to be deployed in large environments. Continuous spatio-temporal maps modelling space and time together suffer from high computational complexity. On the other hand, continuous maps that modelled space and time independently were scalable, computationally efficient and provided good predictions for the topologically uncomplicated area.

The comparison of state-of-the-art methods for robotic mapping showed that proposed criteria derived from the methodology intuitively distinguish optimised human flow models. Moreover, a detailed analysis of service disturbance provides deep insight into the model's strong and weak aspects. We also show that methods specialised in capturing the spatio-temporal distributions of pedestrian flows allow for the planning of trajectories which generate substantially fewer human-robot encounters than time-averaged ones. On testing dataset, the time-sensitive models were able to reduce the number of encounters by order of magnitude if the robot was allowed to drop 10% of the traversals during the busiest times and places and by order of two magnitudes when dropping 60% of traversals.

Part IV

Dataset for Complex Evaluation of Spatio-Temporal Maps

11 Outline

Hypertime, especially regression variant of WHyTeS , proved its usability in human-focused tasks during the COVID-19 outbreak. It was an integral part of a system that predicted the crowdedness of places like shops and pharmacies, which people needed to visit regularly [100]. The idea behind the system lay in the possibility of people taking advantage of it when planning their visits to such places and avoiding long queues, crowded places, and hypothetical exposure to the virus sources - other people.

During that time, a lot of observed places changed their behaviour. For example, as ad hoc government rules changed over time, shops had to change their opening hours or how they provided their services. We wanted our system to react adequately, which led to searching for novelties in a data stream. This idea was first targeted in a bachelor thesis [97] which I supervised. As the COVID-19 outbreak was not problematic to most of the population, the development was abandoned, and this work was not finalised. However, partial results in novelty detection in [97] led to a hypothesis that combining a suitable version of advanced spatial outlier detector [144] with a continuous spatio-temporal model like HyTS could lead to successful novelty detection in human presence and flows. Development and evaluation of such a method need a human-dynamics-centred long-term dataset that includes known outliers of different types. Collecting such a dataset requires a complex setting of an experimental environment, known external influencers of human behaviour, and the possibility to artificially enforce a change in people's behaviour with expected impacts on human flows.

12 Dataset Collection

12.1 Need for Data

During the last decade we have seen much development in the world of robotics, but fully autonomous robotic deployments still remain elusive. One of the primary reasons for this is that while their navigation algorithms work well, their robustness outside of lab conditions can be less-than ideal, and therefore they still require some level of human oversight.

Traditional approaches to robot navigation in uncontrolled environments use static maps combined with reactive planning for unexpected events. However, many unexpected events are caused by human actions, especially movement through the robot's operational environment. Reactive re-planning using sense-plan-act frameworks, such as those used in the Robot Operating System (ROS) [105] *move_base* package, can lead to a perception of clumsiness [106], slow response, and frequent re-planning, resulting in negative emotions towards the robot [23]. Failing to navigate in human-populated or crowded environments is a major obstacle to robotic deployments and their acceptance in the long-term. Therefore, incorporating expected human movement into the robot's navigation and planning is crucial for autonomous robots intended to help people in their environment.

Addressing this, several large projects attempted to work on autonomous robots operating in human-populated environments [9], [101], [102]. These environments tend to be dynamic, with natural daily and seasonal changes [55], but more importantly, their dynamics are shaped by human actions [3]. A promising way to achieve long-term human-aware navigation is to include spatio-temporal

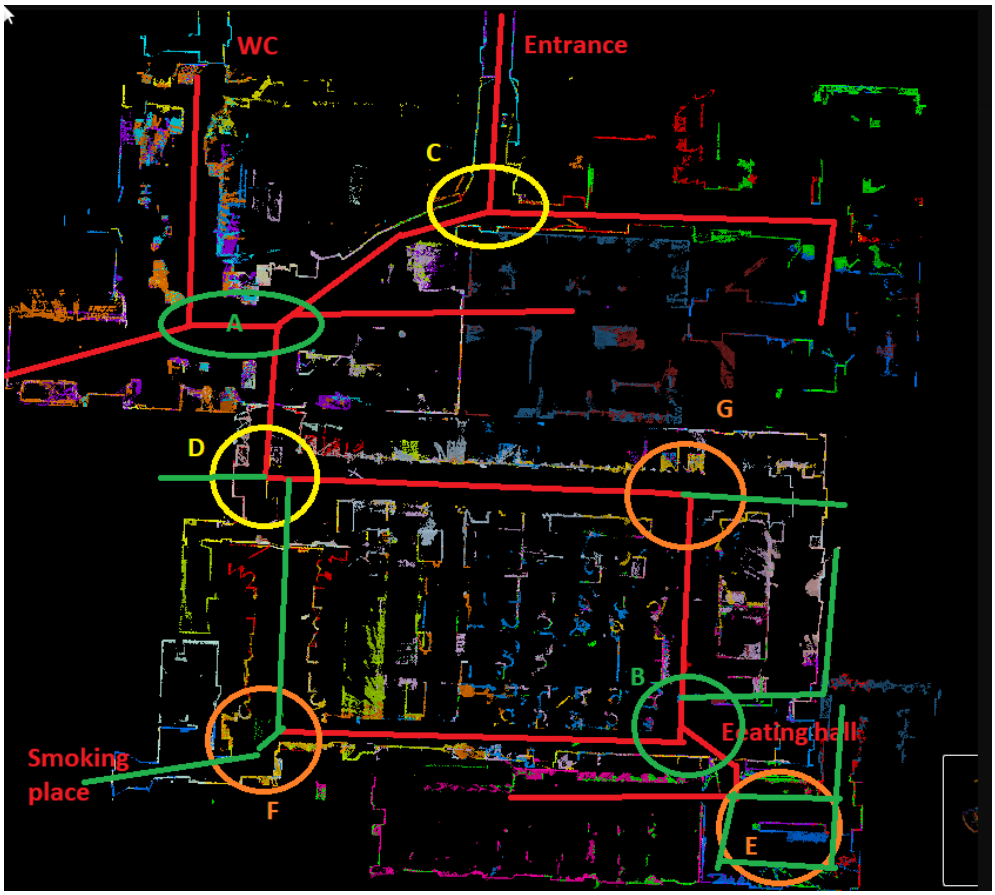

Figure 12.1: Tesla factory lidar scan. Red lines highlight the usual path of people, while green ones are used less frequently. Green circles highlight the most interesting crossroads - A is the nearest crossroad to the bathroom, and B includes a resting area. Yellow circles highlight less frequent crossing with expectedly strong temporal patterns, and orange circles highlight places where people usually stay for some time.

maps into the navigation systems to support decision-making in advance of the navigational task. Robots which forecast the dynamics of human customs and adjust their decisions accordingly perform better at human-centric tasks [9]–[13], [104]. Furthermore, it has been shown that autonomous robots that take human habits into account are more likely to be accepted by society [23], [74], [101].

Developing spatio-temporal predictive maps requires long-term datasets of natural human environments. The most known dataset of this kind is the ATC dataset of a shopping mall [128]. Although it consists of detections taken throughout one year, the data is collected only during opening hours on Wednesdays and Sundays. In addition, researchers from the Lincoln Centre for Autonomous Systems (L-CAS) gathered their own UoL dataset [63] that consists of several consecutive weeks of lidar point clouds and human detections [83]. However, local security rules did not allow researchers to gather the data out of working hours. Using a very similar system, researchers created the UTBM dataset [17] that was supposed to be long-term, similar to the ATC one, and continuous. The effort resulted in a dataset several months long covering entire days. However, due to political problems in the country that led to the universities closing, a large part of the dataset consists of empty halls. Changes in security rules and the following COVID-19 epidemic led to an inability of researchers to continue collecting the data. The meaningful data now includes one month and one week of continuous data.

Despite the discontinuity in the ATC dataset, it is still commonly used [127], [145], [146]. From our experience, a more complex dataset is needed for developing an autonomous system that includes service robots with various tasks. The service robots do not only work during opening hours; they need to cooperate with technical and service staff who work in public places outside of working hours. For example, security robots need to understand differences in behaviour between customers and utility workers. Therefore, we are starting to gather continuous long-term datasets from a factory (see Fig. 12.1). As the data gathering and dataset preparation is planned for the next few years, we are looking for advice and possible cooperation with scientists from the long-term human motion prediction community.

12.2 Datasets

12.2.1 Available Datasets from Other Teams

The ATC dataset [128] covers an area of 900 m^2 over one year. The tasks people are performing are connected with shopping. We can find there different types of behaviour like walking from one shop to another, reading some information, waiting, chatting and so on. However, as was said earlier, the datasets consist only of Thursdays and Sundays and only of working hours. The models developed over such data need not comprise subtle differences between individual working days and Saturday and Sunday. They cannot be tested in an ability to provide apparent differences between days and nights. We had a similar experience working on the UoL dataset [63], Figure 4.10. One can expect that out of the working hours, there is nothing to care much about. However, an autonomous system should be prepared to utilise the time when the halls are empty, and therefore it needs to know what is happening during the night to schedule its tasks accordingly. Moreover, during non-working hours, there is a higher possibility of a security breach [20] that must be distinguished from rare but expectable events or regular utility activities.

One can find complex, human-dynamics-centred datasets like L-CAS [83], THÖR [147], or Magni [148] datasets. Although remarkable, these datasets do not meet our concept of long-term datasets. They do not focus on human customs nor model deterioration.

We usually use open data 'MHT building lecturer office' to test year-long forecasts of our models. The data was gathered during the STRANDS project [9]. They can be found on the project's web pages [149]. The data consists of preprocessed video frames with a frame rate of $0.2Hz$. Every frame consists of a set of depth values captured with a 320×240 RGB-D camera (see Fig. 12.2). The length of the video is more than 2 years with several few-days-long gaps. The video shows one of the lecturer's offices at the University of Lincoln.

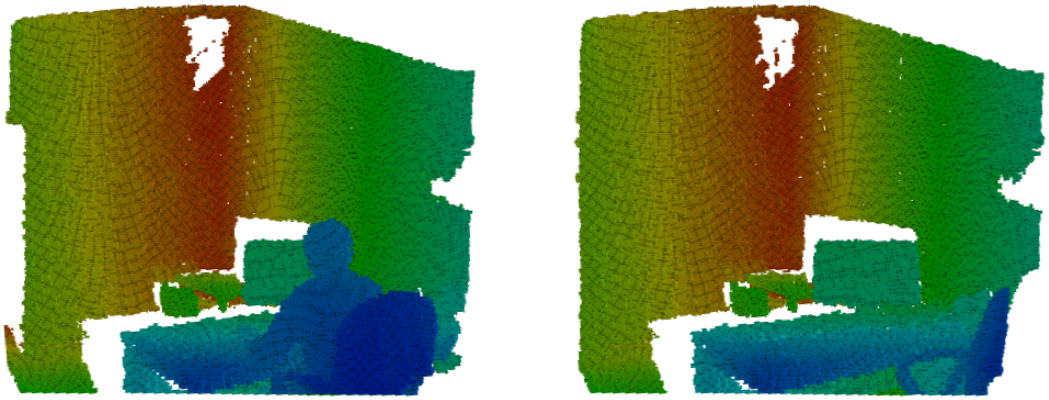

Figure 12.2: MHT building lecturer office dataset: snapshots of the person present and absent.

12.2.2 UTBM Dataset

The second dataset we usually use to test our models is the UTBM dataset [17], which we collected during cooperation with Distributed Artificial Intelligence and Knowledge Laboratory (CIAD) at Université de technologie de Belfort Montbéliard. The data is not directly accessible to the public but only by request due to privacy concerns. It contains two long-term segments gathered during one year. The first segment was collected in one of the UTBM building halls of approximately $500m^2$. As shown in Fig. 9.4, a Velodyne 32-layer lidar was placed in the reception near the building door to ensure safe 24h operation. The spatial placement of the lidar was carefully determined to ensure maximum field-of-view (i.e., approximately $200m^2$) of the hall beyond the glass windows. The raw data of the lidar was recorded to ROS rosbags 24 hours per day for several months. However, due to political unrest, only one month (March 2019) includes normal student behaviour. The second segment, one week long, was collected in December 2019 in a similar way. There is also the third segment, still in the post-processing stage, collected in the same hall but with a 128-layer lidar outside but close to the reception. The data collection was performed for three consecutive days during March 2023.

12.2.3 Dataset under Construction

The MHT dataset is two years long but includes only one significant phenomenon connected to human customs. The dataset's length allows us to test a model's deterioration over months-long periods. However, testing the models on one phenomenon can lead to overfitting the chosen scenario. The UTBM dataset consists of human flows in the entrance hall, but it is limited only to a few weeks. Although the data consists of many different people, their usual behaviour is walking through or waiting in the hall, with rare exceptions of utility workers providing their services. Moreover, the usable data cover only one topologically trivial area - the hall. Our experiences and needs led us to initiation of a new dataset collection.

We made an agreement with the management and owners of a factory that we have the freedom to place sensors usually used in mobile robotics. The factory consists of two large sheds and multiple rooms with an overall area of approximately $2500m^2$, more than twice the area of the ATC dataset. Figure 12.1 shows a cross-section of a lidar scan of the factory. Our team, in cooperation with the factory management, chose and highlighted the best areas for data collection. The lines highlight the frequent paths the employees usually use. The circles highlight important crossings and places where employees assemble.

We aim to gather a new, similarly long or longer dataset than the MHT building lecturer office composed of people working on multiple tasks in an environment with a more complicated structure than the UTBM dataset. We expect people to perform more types of tasks than in UTBM datasets. Besides walking, waiting, and cleaning the area, we expect activities like resting, eating, refilling their water containers, and, similarly to the MHT dataset, working in front of their tables. That can give us a more generalised perspective in analysing the deterioration of models over time and the rules of pedestrian flows. Contrary to the ATC dataset, we also plan to gather the data continually during all days of the week and outside the usual working hours.

The number of people present in the area will probably not exceed 50 people, which is less than in UTBM and ATC datasets. However, two companies are in the building fabricating different products while sharing changing rooms,

bathrooms, and the official entrance. One company has been present there for approximately three decades, while the second one is starting up, hiring new employees, and preparing new operations. The building is positioned between two cities, so a non-public bus for employees connects with public transportation. Such transportation is felicitous because the entry time of a large section of employees is determined by the most prolonged delay in documented public transit. The delays are usually connected with weather changes that can be included in the models and allow researchers to interconnect the predictions of human customs with weather forecasting.

The management is in favour of the experiments. They see cooperation with researchers as an opportunity to raise the working morale and moderate discomfort due to experiments as an acceptable price. On the other hand, the workers positively perceive the experiments as a non-traditional event in an overall invariable work. As a result, it is possible to agree with the management to delay the bus or move a shift a little bit to test the ability of models to detect unexpected or anomalous events as a part of the experimental design. There is also a possibility to perform field experiments with piloted or supervised robots.

12.2.4 Challenges in Data Gathering

The factory is a harsh environment. There is heavy machinery creating strong magnetic fields. The data cables need to be carried through high-voltage pipes. Computers need to be placed far away from heavy machinery and reactive chemicals. Nothing can be placed near the gas tubes or hydrogen tanks. The environment is dusty, and the air temperature in some places exceeds 40 degrees Celsius. Multiple security fences around some machines block WiFi communication between the computers. On the other hand, the ceilings are $4 - 8$ meters high, which can give sensors the required scope.

Unlike in the UTBM dataset, human flows are not the only dynamics in the environment. There are also chairs moving from one place to another. Some cupboards are used, and some working positions are occupied only on different days. Large machines are run occasionally, and some rooms are accessible only for one or a few specific people during their particular tasks. Those dynamics

differ from human flows but are tightly bound to people's activity. Therefore, the structure of the area is not purely static but "reacts" and changes its shape according to the actual organisation of the work. A successful autonomous system should predict these small structural changes because they are connected with the differences in human flows.

12.2.5 Privacy Issues

The privacy and data protection rules limit personal or otherwise sensitive data manipulation only inside the building. Although the management does not plan to use the personal data from the measurements, they find the sensors' placement as motivational components for the workers. The workers are used to working with expensive products, and their only concern about the data gathering is not to be identified by the sensors when they smoke near the hydrogen tanks.

The internet connection to the factory is stable but slow for data transfer. The data must be transferred from the factory manually, but the computers and sensors can be handled using a secure connection from outside. Personal or sensitive data needs to be processed or anonymised inside the compound.

12.3 Data Collection

Our primary way to collect data of human motion tends to be with lidar sensors. These have the advantage of being small, light-weight, and not intrusive on the environment. They also yield more fruitful data than highly bespoke sensors such as door position detectors, and detecting people passing through an area with a single beam. Furthermore, compared to cameras, lidar data is much easier to process as properties such as depth can be directly estimated (especially at long range), they're insensitive to lighting conditions, whilst also manipulation of raw data does not open privacy issues. Moreover, they have a large field-of-view. However this raises the question of whether the community would agree with these assessments for the task of human detection, or perhaps whether it would be advantageous to include other sensors in the data collection process.

The current arrangement for data collection involves having a series of com-

puters set up running ROS. Each of these is running the necessary drivers of the lidar sensor, in this case a 128-layer Ouster sensor giving 3D coverage of the environment at $10Hz$. The sensors have slightly overlapping fields-of-view, to ensure the consistent ability to track individuals. The sensor is left running continuously, with another ROS node collecting the data, compressing it, and saving it to disk.

The throughput of the raw lidar data amounts to just over 100MB per second per sensor, an unwieldy size for this kind of usage. Consequently, the run-time compression of the data is essential. While lossless lz4 compression offered approximately a 50% saving compared to the raw data, more than this kind of compression is needed for long-term use. As a result, a bespoke lossy compression was added.

The compression algorithm works by taking a key-frame from the lidar, and then in subsequent frames only storing lidar points with significant changes in their spatial position. These are found by matching the corresponding lidar beams between the frames, and finding the euclidean change in position. Jitter inevitably is present in the data, so there is some thresholding, but the suppression of all static background lidar points is not a significant challenge. Furthermore, when storing only key-frames and the deltas, traditional compression algorithms can still be used on the data to reduce the file-size further. As a result, the volume of data being stored was able to be reduced to below 0.1% of the original file size.

The data that saved as a result of this process has been modified from the raw data. Visibly, the jitter present on the walls of rooms is completely absent in the data processed in this way, while people are still visible moving around the environment. While very effective in making the dataset a manageable size, it is not clear whether the community would appreciate the data being processed in such a way. While generally the raw data has the benefit of being closer to reality, some compromise needs to be found, so the question exists if this kind of compression is acceptable for the community. We believe that such an approach will not affect the quality or usability of the dataset, as the dynamic objects will be captured in full, real-time detail, while static background objects such as walls, less likely to be of interest, will still be captured, but less often.

The rest of the data collection pipeline is run offline. The compressed data can be downloaded from each of the machines periodically, and then the data

can be worked on. Firstly, they are decompressed individually back into a usable form. As mentioned, the background lidar points will be completely static, but any dynamic areas of the pointcloud will be present at the full frame rate.

Then, these dynamic points are passed to human detection algorithms to extract the semantic data from the scene. The detection algorithms are those used in the aforementioned UTBM data collection [83]. Specifically, the Adaptive Clustering algorithm is implemented on each frame of the pointcloud to divide it into different segments, ideally with each segment corresponds to an object. The segmentation performance of the algorithm for pedestrians is considered to be state-of-the-art [150]. Then the segments are filtered using a human volumetric model. On the other hand, since the UTBM building is completely closed at night and ensures that no one is stranded inside, we use the evening data as a reference background and then match it with daytime data to remove any false positive samples efficiently. As a result, human samples can be easily extracted from the point cloud without using any machine learning methods. It is worth noting that, there is an assumption here that the moving objects in the UTBM building fitted in the human volume model are all human beings.

Once the raw human detections have been extracted from the dataset, a final stage of filtering is applied before the final data is ready. This filtering process removes any spurious detections of people from the scene, and fills in any missing detections where otherwise there was continuous detections. As the raw data was collected at 10Hz, data association for tracking is relatively straight-forward, however as the original data is present, it is possible to verify this manually, or apply different methods for human detection and tracking.

The data about human detections and tracking is therefore provided as a convenience for people wishing to work with the dataset, however there is still the option for people to try their own detection and tracking algorithms on the data should they wish to.

13 Summary

Nowadays, we are collecting exceptional, long-term dataset of human flows for the complex evaluation of robotic spatio-temporal maps and the development of their new capabilities. We have looked at why long-term datasets of people will be important for the future of long-term autonomous robotics. Currently, there is a lack of high-quality datasets available for this task. In an attempt to correct this, we have started collecting our own dataset based on 3D lidars, to be used for analysis of the typical presence and movement patterns of people, and human-induced dynamics in their working environment. The dataset details, along with the forms to provide potential feedback and suggestions, is available at `http://3l4ar.science`.

Part V

Conclusion

Summary of thesis

In recent years, the main objective of the papers that replace stationary probability with a Fourier transform-based time series modelling of events in robotic maps was to persuade the readers that using periodicities is beneficial for the autonomous planning of different robotic tasks. Although the benefit was evident in the long-term deployment of robots in a human society [9], the theoretical results in the experiments were, sometimes, not too sound. For example, in [16], the authors compare the quality of predictions of FreMEn and HyT with a classical occupancy grid. The error reduction of FreMEn's model was between 0 and 0.1% while HyT was better by 1% - in a chosen spatio-temporal scenario and selected parameters. Moreover, the model that predicts 0 everywhere was omitted, although we can expect it to perform much better; see Table 9.4. The reason was studied in [26] and discussed in [23] - when a human is understood as a point in space and time, the event, human detected at coordinate (x, y, t), is very rare compared to a large number of cells in $4D$ spatio-temporal map. As robotics is very loosely connected to social studies, it was also sometimes hard to persuade the community that it is not such an imprudent idea to model human behaviour as a repetitive process with negligible progress.

We proposed a methodology for the quality of spatio-temporal maps comparison [17]. Using this methodology, we derived a new type of criterion, Expected Encounters [23], that evaluates the quality of these maps by verifying the ability of an autonomous system to plan and schedule its navigational tasks when using their forecasts. Eventually, we showed the superiority of FreMEn-derived, especially continuous methods, and supported that with a field experiment. As of now, we believe that the question 'whether the modelling of periodical events in human-populated areas is beneficial for deployed autonomous robots' is successfully answered.

We developed a continuous version of FreMEn, denoted as HyTS. We proved it is converging to a helpful model faster than FreMEn [18], and its size is by two orders of magnitude smaller than its discrete competitors [20], [21], FreMEn and STeF-Map. During the COVID-19 outbreak, we developed a computationally simplified variant of HyTS denoted as WHyTeS, which success-

fully served in the production version of our country-wide application `https://kdynakoupit.cz` [100]. This application estimates the crowdedness of user-defined important places, with the original purpose of helping people to evade crowds in shops and pharmacies when meeting other people led to a rise in the probability of unfamiliar disease spread. Despite the successes of existing methods, developing continuous spatio-temporal maps should start afresh considering the proposed methodology and the possibility of defining new specialised criteria. I believe that the comprehensive overview and comparison of various approaches to spatio-temporal modelling [17] could serve as a good starting point. However, the development also needs specialised, high-quality datasets. One of such datasets we already started to collect [24].

Philosophically, we claim [17] that the human-populated environment does not contain only spatial structures like walls but also temporal structures derived from human routines. Those routines are based on work management, directly connected to the human calendar and clock. Similarly to any worker, autonomous robots need to be aware of other workers' routines and customs to avoid spoiling the psychosocial atmosphere. This "right time to serve" concept constitutes a novel paradigm in robotic autonomy, studied, besides other things related to the robotic perception of time, by a new robotic field Chronorobotics [70], [109].

Fulfilment of targets

Although the goal to create "a general tool to model spatio-temporal phenomena useful for robots operating in large, human-populated areas for long periods" is megalomaniacal, reflecting my youthful imprudence, it can be, with some level of leniency, proclaimed as achieved. The proposed method and its variants were applied to a diverse real-world dataset and successfully compared with state-of-the-art methods [21], [23]. It proved its scalability when employed in an application covering a country-wide area [100], and it successfully targeted an unheeded quality of 'master-servant' interaction - the acceptance of provided service [17].

A better acceptance by human society of autonomous mobile robot that uses spatio-temporal map for scheduling their task compared to the opposite was proved in a field experiment with industrial-grade humanoid [17], [23]. Although

human subjects of the experiment belonged to an academic milieu, they were not warned in advance, did not know they were subjects of study, and did not understand the purpose of the experiment. The difference in human subjects' reactions surpasses our expectations.

The original specific goals were declared when we expected to work with data that included occurrences and non-occurrences of the studied phenomenon [16]. However, for higher dimensional tasks like modelling the presence and velocity of people together, the inclusion of non-occurrences in the data was computationally infeasible [17], [21], [23]. Therefore, the development of HyTS changed its expected direction, and, for example, further development of multiclass classification was omitted in my work. Other specific goals were successfully targeted: we applied HyT to binary, door state [16] and person presence classification and novelty detection [18], and HyTS to human frequency analysis [16], [20], regression describing the crowdedness of specific places [100], and novelty detection in time series [97]. Spatio-temporal novelty detection was not defined nor evaluated. However, we are currently collecting a dataset that, as we expect, could serve this purpose [24]. Above the original expectations, HyTS was successfully applied to modelling the directions and velocities of human flows [17], [21], [23], which was crowned with developing a methodology to compare miscellaneous spatio-temporal representations [17].

Regarding the collection of the datasets, we collected months-long spatial and spatio-directional human detections on a 24/7 basis [17]. However, for political reasons, we possess only five weeks of a continuous window, with another one-week window after eight months. Although extraordinary in many ways, those data are not accessible to the general public and can be handed over only on demand due to privacy concerns. As mentioned above, we currently collect new dataset on a 24/7 basis, which is planned to be public from the early stages of collection [24]. Besides the human-induced datasets, our laboratory collected navigation maps covering natural changes in an outdoor environment [96], [151], [152]. My assistance was very sporadic, with only one exception [99], where I directed the efforts and actively cooperated on such a data collection.

List of candidate's work related to the thesis

Journals (Impact)

- T. Vintr, J. Blaha, M. Rektoris, *et al.*, "Toward benchmarking of long-term spatio-temporal maps of pedestrian flows for human-aware navigation", *Frontiers in Robotics and AI*, vol. 9, p. 890 013, 2022 (2 known citations)

- T. Krajník, T. Vintr, S. Molina, *et al.*, "Warped hypertime representations for long-term autonomy of mobile robots", *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3310–3317, 2019 (30 known citations)

Journals (Reviewed)

- T. Vintr, K. Eyisoy, and T. Krajník, "A practical representation of time for the human behaviour modelling", *Forum Statisticum Slovacum*, vol. 14, pp. 61–75, 2018, ISSN: 1336-7420 (unknown number of citations)

Conferences (WoS)

- T. Vintr, Z. Yan, K. Eyisoy, *et al.*, "Natural criteria for comparison of pedestrian flow forecasting models", in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2020, pp. 11 197–11 204 (8 known citations)

- T. Vintr, S. Molina, R. Senanayake, *et al.*, "Time-varying pedestrian flow models for service robots", in *2019 European Conference on Mobile Robots (ECMR)*, IEEE, 2019, pp. 1–7 (25 known citations)

- T. Vintr, Z. Yan, T. Duckett, *et al.*, "Spatio-temporal representation for long-term anticipation of human presence in service robotics", in *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, 2019, pp. 2620–2626 (25 known citations)

- T. Vintr, K. Eyisoy, V. Vintrová, *et al.*, "Spatiotemporal models of human activity for robotic patrolling", in *International Conference on Modelling and Simulation for Autonomous Systems*, Springer, 2018, pp. 54–64 (11 known citations)

Other Publications

- T. Vintr, G. Broughton, T. Roucek, *et al.*, "Dataset collection for long-term forecasting of human presence, motion and activity", in *ICRA 2023 Workshop on Long-term Human Motion Prediction*, 2023 (0 known citations)

- T. Krajník, T. Vintr, G. Broughton, *et al.*, "Chronorobotics: Representing the structure of time for service robots", in *Proceedings of the 2020 4th International Symposium on Computer Science and Intelligent Control*, 2020, pp. 1–8 (6 known citations)

- T. Vintr, S. Molina, R. Senanayake, *et al.*, "Spatio-temporal representation of time-varying pedestrian flows", in *ICRA 2019 Workshop on Long-term Human Motion Prediction*, 2019 (1 known citation)

- T. Vintr, S. Molina Mellado, G. Cielniak, *et al.*, "Spatiotemporal models for motion planning in human populated environments", 2017 (8 known citations)

- T. Krajnık, M. Hanheide, T. Vintr, *et al.*, "Towards automated benchmarking of robotic experiments", in *ICRA Workshop on Reproducible Research in Robotics*, 2017 (3 known citations)

Applied results

- Laboratory of Chronorobotics, *Kdynakoupit*, `https://kdynakoupit.cz/`, Accessed: 2022-02-02, 2019 (unknown number of citations)

List of candidate's work non-related to the thesis

Journals (Impact)

- Z. Rozsypálek, T. Rouček, T. Vintr, *et al.*, "Multidimensional particle filter for long-term visual teach and repeat in changing environments", *IEEE Robotics and Automation Letters*, vol. 8, no. 4, pp. 1951–1958, 2023 (2 known citations)

- G. Broughton, J. Janota, J. Blaha, *et al.*, "Embedding weather simulation in auto-labelling pipelines improves vehicle detection in adverse conditions", *Sensors*, vol. 22, no. 22, p. 8855, 2022 (6 known citations)

- T. Vintr, V. Vintrova, and H. Rezankova, "Poisson distribution based initialization for fuzzy clustering", *Neural Network World*, vol. 22, no. 2, p. 139, 2012 (3 known citations)

Conferences (WoS)

- G. Broughton, P. Linder, T. Rouček, *et al.*, "Robust image alignment for outdoor teach-and-repeat navigation", in *2021 European Conference on Mobile Robots (ECMR)*, IEEE, 2021, pp. 1–6 (3 known citations)

- L. Halodová, E. Dvořráková, F. Majer, *et al.*, "Predictive and adaptive maps for long-term visual navigation in changing environments", in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2019, pp. 7033–7039 (21 known citations)

- F. Majer, L. Halodová, T. Vintr, *et al.*, "A versatile visual navigation system for autonomous vehicles", in *International Conference on Modelling and Simulation for Autonomous Systems*, Springer, 2018, pp. 90–110 (9 known citations)

- L. Halodová, E. Dvořáková, F. Majer, *et al.*, "Adaptive image processing methods for outdoor autonomous vehicles", in *International Confer-*

ence on Modelling and Simulation for Autonomous Systems, Springer, 2018, pp. 456–476 (19 known citations)

- T. Krajník, F. Majer, L. Halodová, *et al.*, "Navigation without localisation: Reliable teach and repeat based on the convergence theorem", in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2018, pp. 1657–1664 (46 known citations)

- T. Vintr, L. Pastorek, and H. Rezankova, "Autonomous robot navigation based on clustering across images", in *Research and Education in Robotics-EUROBOT 2011: International Conference, Prague, Czech Republic, June 15-17, 2011. Proceedings*, Springer, 2011, pp. 310–320 (5 known citations)

Bibliography

[1] F. Dayoub, T. Duckett, G. Cielniak, *et al.*, "An adaptive spherical view representation for navigation in changing environments", in *European Conference on Mobile Robots-ECMR 2009*, 2009.

[2] P. Biber and T. Duckett, "Experimental analysis of sample-based maps for long-term slam", *The International Journal of Robotics Research*, vol. 28, no. 1, pp. 20–33, 2009.

[3] G. D. Tipaldi, D. Meyer-Delius, and W. Burgard, "Lifelong localization in changing environments", *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1662–1678, 2013.

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

[5] T. Krajník, J. P. Fentanes, J. M. Santos, and T. Duckett, "Fremen: Frequency map enhancement for long-term mobile robot autonomy in changing environments", *IEEE Transactions on Robotics*, vol. 33, no. 4, pp. 964–977, 2017.

[6] W. Churchill and P. Newman, "Experience-based navigation for long-term localisation", *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1645–1661, 2013.

[7] K. Konolige and J. Bowman, "Towards lifelong visual maps", in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, IEEE, 2009, pp. 1156–1163.

[8] S. Hochdorfer and C. Schlegel, "Towards a robust visual slam approach: Addressing the challenge of life-long operation", in *Advanced Robotics, 2009. ICAR 2009. International Conference on*, IEEE, 2009, pp. 1–6.

[9] N. Hawes, C. Burbridge, F. Jovan, *et al.*, "The strands project: Long-term autonomy in everyday environments", *IEEE Robotics & Automation Magazine*, vol. 24, no. 3, pp. 146–156, 2017.

[10] J. M. Santos, T. Krajník, J. P. Fentanes, and T. Duckett, "Lifelong information-driven exploration to complete and refine 4-d spatio-temporal maps", *IEEE Robotics and Automation Letters*, vol. 1, no. 2, pp. 684–691, 2016.

[11] L. Kunze, N. Hawes, T. Duckett, M. Hanheide, and T. Krajník, "Artificial intelligence for long-term robot autonomy: A survey", *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4023–4030, 2018.

[12] J. M. Santos, T. Krajník, and T. Duckett, "Spatio-temporal exploration strategies for long-term autonomy of mobile robots", *Robotics and Autonomous Systems*, vol. 88, pp. 116–126, 2017.

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

[14] T. Krajník, M. Kulich, L. Mudrová, R. Ambrus, and T. Duckett, "Where's waldo at time t? using spatio-temporal models for mobile robot search", in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2015, pp. 2140–2146.

[15] T. Krajník, J. P. Fentanes, O. M. Mozos, T. Duckett, J. Ekekrantz, and M. Hanheide, "Long-term topological localisation for service robots in dynamic environments using spectral maps", in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2014, pp. 4537–4542.

[16] T. Krajník, T. Vintr, S. Molina, *et al.*, "Warped hypertime representations for long-term autonomy of mobile robots", *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3310–3317, 2019.

[17] T. Vintr, J. Blaha, M. Rektoris, *et al.*, "Toward benchmarking of long-term spatio-temporal maps of pedestrian flows for human-aware navigation", *Frontiers in robotics and AI*, vol. 9, 2022.

[18] T. Vintr, K. Eyisoy, V. Vintrová, Z. Yan, Y. Ruichek, and T. Krajník, "Spatiotemporal models of human activity for robotic patrolling", in *International Conference on Modelling and Simulation for Autonomous Systesm*, Springer, 2018, pp. 54–64.

[19] T. Vintr, K. Eyisoy, and T. Krajník, "A practical representation of time for the human behaviour modelling", *Forum Statisticum Slovacum*, vol. 14, pp. 61–75, 2018, ISSN: 1336-7420.

[20] T. Vintr, Z. Yan, T. Duckett, and T. Krajník, "Spatio-temporal representation for long-term anticipation of human presence in service robotics", in *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, 2019, pp. 2620–2626.

[21] T. Vintr, S. Molina, R. Senanayake, *et al.*, "Time-varying pedestrian flow models for service robots", in *2019 European Conference on Mobile Robots (ECMR)*, IEEE, 2019, pp. 1–7.

[22] T. Vintr, J. Blaha, J. Ulrich, *et al.*, "Estimating the dominant frequencies of real-valued cyclic processes in natural environments for long-term operation of autonomous robots", *Robotics and Autonomous Systems*, In review.

[23] T. Vintr, Z. Yan, K. Eyisoy, *et al.*, "Natural criteria for comparison of pedestrian flow forecasting models", in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2020, pp. 11 197–11 204.

[24] T. Vintr, G. Broughton, T. Roucek, Z. Yan, and T. Krajnık, "Dataset collection for long-term forecasting of human presence, motion and activity", 2023.

[25] T. P. Kucner, M. Magnusson, S. Mghames, *et al.*, "Survey of maps of dynamics for mobile robots", *The International Journal of Robotics Research*, vol. 0, no. 0, p. 02 783 649 231 190 428, 0. DOI: 10.1177/02783649231190428. eprint: https://doi.org/10.1177/02783649231190428. [Online]. Available: https://doi.org/10.1177/02783649231190428.

[26] F. Kubiš, "Application of spatiotemporal modeling used in robotics for demand forecast", B.S. thesis, Czech Technical University in Prague, 2020.

[27] S. T. O'Callaghan and F. T. Ramos, "Gaussian process occupancy maps", *The International Journal of Robotics Research*, vol. 31, no. 1, pp. 42–62, 2012.

[28] C. Cadena, L. Carlone, H. Carrillo, *et al.*, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age", *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[29] D. Austin, L. Fletcher, and A. Zelinsky, "Mobile robotics in the long term-exploring the fourth dimension", in *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)*, IEEE, vol. 2, 2001, pp. 613–618.

[30] G. D. Finlayson and S. D. Hordley, "Color constancy at a pixel", *JOSA A*, vol. 18, no. 2, pp. 253–264, 2001.

[31] J. Mount and M. Milford, "2d visual place recognition for domestic service robots at night", in *2016 IEEE international conference on robotics and automation (ICRA)*, IEEE, 2016, pp. 4822–4829.

[32] A. Gawel, T. Cieslewski, R. Dubé, M. Bosse, R. Siegwart, and J. Nieto, "Structure-based vision-laser matching", in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2016, pp. 182–188.

[33] D. M. Rosen, J. Mason, and J. J. Leonard, "Towards lifelong feature-based mapping in semi-static environments", in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*, IEEE, 2016, pp. 1063–1070.

[34] S. Lowry, G. Wyeth, and M. Milford, "Unsupervised online learning of condition-invariant images for place recognition", in *Proc. Australasian Conf. on Robot. and Automation*, Citeseer, 2014.

[35] N. Sünderhauf, P. Neubert, and P. Protzel, "Predicting the change–a step towards life-long operation in everyday environments", *Robotics Challenges and Vision (RCV2013)*, p. 17, 2014.

[36] T. Krajník, *Long–Term Autonomy of Mobile Robots in Changing Environments*. CTU Press, 2018.

[37] K. Van Laerhoven, D. Kilian, and B. Schiele, "Using rhythm awareness in long-term activity recognition", in *12th IEEE International Symposium on Wearable Computers(ISWC)*, IEEE, 2008, pp. 63–66.

[38] U. Blanke and B. Schiele, "Daily routine recognition through activity spotting", in *International Symposium on Location-and Context-Awareness*, Springer, 2009, pp. 192–206.

[39] Z. Zhou, D. S. Matteson, D. B. Woodard, S. G. Henderson, and A. C. Micheas, "A spatio-temporal point process model for ambulance demand", *Journal of the American Statistical Association*, vol. 110, no. 509, pp. 6–15, 2015.

[40] V. H. Bennetts, K. Kamarudin, T. Wiedemann, T. P. Kucner, S. L. Somisetty, and A. J. Lilienthal, "Multi-domain airflow modeling and ventilation characterization using mobile robots, stationary sensors and machine learning", *Sensors (Basel, Switzerland)*, vol. 19, no. 5, 2019.

[41] F. L. Bayisa, M. Ådahl, P. Rydén, and O. Cronie, "Large-scale modelling and forecasting of ambulance calls in northern sweden using spatio-temporal log-gaussian cox processes", *Spatial Statistics*, vol. 39, p. 100 471, 2020.

[42] R. Senanayake, O. Simon Timothy, and F. Ramos, "Predicting spatio-temporal propagation of seasonal influenza using variational gaussian process regression.", in *AAAI*, 2016, pp. 3901–3907.

[43] Z. Zhou and D. S. Matteson, "Predicting melbourne ambulance demand using kernel warping", *The Annals of Applied Statistics*, vol. 10, no. 4, pp. 1977–1996, 2016.

[44] Ç. Ak, Ö. Ergönül, İ. Şencan, M. A. Torunoğlu, and M. Gönen, "Spatiotemporal prediction of infectious diseases using structured gaussian processes with application to crimean–congo hemorrhagic fever", *PLoS neglected tropical diseases*, vol. 12, no. 8, e0006737, 2018.

[45] S. H. Garrido Mejía *et al.*, "Predicting crime in bogota using kernel warping", M.S. thesis, Uniandes, 2018.

[46] J. S. M. Pabón, M. D. Rubio, Y. Castaño, A. J. Riascos, and P. R. Díaz, "A manifold learning data enrichment methodology for homicide prediction", in *2020 7th International Conference on Behavioural and Social Computing (BESC)*, IEEE, 2020, pp. 1–4.

[47] Z. Zhou and D. S. Matteson, "Predicting ambulance demand: A spatio-temporal kernel approach", in *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 2015, pp. 2297–2303.

[48] A. Gilardi, R. Borgoni, and J. Mateu, "A spatio-temporal model for events on road networks: An application to ambulance interventions in milan", *Preface XIX 1 Plenary Sessions*, p. 702, 2021.

[49] S. Nilsang and C. Yuangyai, "Activity detection for multi-factors of ambulance demand areas: A case study in bangkok", in *AIP Conference Proceedings*, AIP Publishing LLC, vol. 2397, 2021, p. 020 001.

[50] V. C. Guizilini and F. T. Ramos, "A nonparametric online model for air quality prediction.", in *AAAI*, 2015, pp. 651–657.

[51] 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.

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

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

[54] A. Tompkins and F. Ramos, "Periodic kernel approximation by index set fourier series features", in *Uncertainty in Artificial Intelligence*, PMLR, 2020, pp. 486–496.

[55] T. Krajnik, J. P. Fentanes, G. Cielniak, C. Dondrup, and T. Duckett, "Spectral analysis for long-term robotic mapping", in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2014, pp. 3706–3711.

[56] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees", *Autonomous Robots*, 2013, Software available at `http://octomap.github.com`. DOI: `10.1007/s10514-012-9321-0`. [Online]. Available: `http://octomap.github.com`.

[57] T. Krajnik, J. Santos, B. Seemann, and T. Duckett, "Froctomap: An efficient spatio-temporal environment representation", *Advances in Autonomous Robotics Systems*, p. 269, 2014. DOI: `10.1007/978-3-319-10401-0.`.

[58] T. Krajník, J. P. Fentanes, J. Santos, K. Kusumam, and T. Duckett, "FreMEn: Frequency map enhancement for long-term
mobile robot autonomy in changing environments", in *ICRA Workshop on Visual Localization in Changing Environments*, 2015.

[59] S. Bagchi and S. K. Mitra, "The nonuniform discrete fourier transform", in *Nonuniform Sampling: Theory and Practice*, F. Marvasti, Ed. Boston, MA: Springer US, 2001, pp. 325–360, ISBN: 978-1-4615-1229-5. DOI: `10.1007/978-1-4615-1229-5\_7`. [Online]. Available: `https://doi.org/10.1007/978-1-4615-1229-5\_7`.

[60] J. P. Fentanes, B. Lacerda, T. Krajník, N. Hawes, and M. Hanheide, "Now or later? predicting and maximising success of navigation actions from

long-term experience", in *2015 IEEE international conference on robotics and automation (ICRA)*, IEEE, 2015, pp. 1112–1117.

[61] T. Krajník, M. Kulich, L. Mudrová, R. Ambrus, and T. Duckett, "Where's waldo at time t? using spatio-temporal models for mobile robot search", in *IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2015, pp. 2140–2146.

[62] F. Jovan, J. Wyatt, N. Hawes, and T. Krajník, "A poisson-spectral model for modelling temporal patterns in human data observed by a robot", in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2016, pp. 4013–4018.

[63] S. Molina, G. Cielniak, T. Krajník, and T. Duckett, "Modelling and predicting rhythmic flow patterns in dynamic environments", in *Annual Conference Towards Autonomous Robotic Systems*, Springer, 2018, pp. 135–146.

[64] S. Molina, G. Cielniak, and T. Duckett, "Go with the flow: Exploration and mapping of pedestrian flow patterns from partial observations", in *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, 2019, pp. 9725–9731.

[65] F. Surma, T. P. Kucner, and M. Mansouri, "Multiple robots avoid humans to get the jobs done: An approach to human-aware task allocation", in *2021 European Conference on Mobile Robots (ECMR)*, IEEE, 2021, pp. 1–6.

[66] T. Krajník, J. M. Santos, and T. Duckett, "Life-long spatio-temporal exploration of dynamic environments", in *2015 European Conference on Mobile Robots (ECMR)*, IEEE, 2015, pp. 1–8.

[67] M. Kulich, T. Krajník, L. Přeučil, and T. Duckett, "To explore or to exploit? learning humans' behaviour to maximize interactions with them", in *International Workshop on Modelling and Simulation for Autonomous Systems*, Springer, 2016, pp. 48–63.

[68] M. Hanheide, D. Hebesberger, and T. Krajník, "The when, where, and how: An adaptive robotic info-terminal for care home residents", in *ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17, New York, NY, USA: ACM, 2017, pp. 341–349, ISBN: 978-1-4503-4336-7. DOI: 10.1145/2909824.3020228. [Online]. Available: http://doi.acm.org/10.1145/2909824.3020228.

[69] T. Krajník, *Fremen [source code].* https://github.com/gestom/fremen, 2017.

[70] T. Krajník, *Spatio-temporal representations for long-term mobile robot autonomy*, oral presentation, Habilitation lecture at CTU, Prague, 2019.

[71] A. Elfes, "Using occupancy grids for mobile robot perception and navigation", *Computer*, vol. 22, no. 6, pp. 46–57, 1989.

[72] Z. Wang, R. Ambrus, P. Jensfelt, and J. Folkesson, "Modeling motion patterns of dynamic objectsby iohmm", in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014), 14-18 Sept. 2014, Chicago, IL*, IEEE conference proceedings, 2014, pp. 1832–1838.

[73] T. P. Kucner, M. Magnusson, E. Schaffernicht, V. H. Bennetts, and A. J. Lilienthal, "Enabling flow awareness for mobile robots in partially observable environments", *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 1093–1100, 2017.

[74] L. Palmieri, T. P. Kucner, M. Magnusson, A. J. Lilienthal, and K. O. Arras, "Kinodynamic motion planning on gaussian mixture fields", in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2017, pp. 6176–6181.

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

[76] T. Vintr, S. Molina Mellado, G. Cielniak, T. Duckett, T. Krajnik, *et al.*, "Spatiotemporal models for motion planning in human populated environments", *lincol.ac.uk*, 2017.

[77] P. G. Gould, A. B. Koehler, J. K. Ord, R. D. Snyder, R. J. Hyndman, and F. Vahid-Araghi, "Forecasting time series with multiple seasonal patterns", *European Journal of Operational Research*, vol. 191, no. 1, pp. 207–222, 2008.

[78] P. J. Brockwell and R. A. Davis, *Introduction to time series and forecasting*. springer, 2016.

[79] L. Xu and M. I. Jordan, "On convergence properties of the em algorithm for gaussian mixtures", *Neural computation*, vol. 8, no. 1, pp. 129–151, 1996.

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

[81] A. H. Barnett, J. Magland, and L. af Klinteberg, "A parallel nonuniform fast fourier transform library based on an "exponential of semicircle" kernel", *SIAM Journal on Scientific Computing*, vol. 41, no. 5, pp. C479–C504, 2019.

[82] A. H. Barnett, "Aliasing error of the exp(β1- z2) kernel in the nonuniform fast fourier transform", *Applied and Computational Harmonic Analysis*, vol. 51, pp. 1–16, 2021.

[83] H. Yan, Z. Zhang, and J. Zou, "An online spatio-temporal model for inference and predictions of taxi demand", in *2017 IEEE International Conference on Big Data (Big Data)*, IEEE, 2017, pp. 3550–3557.

[84] R. Ortiz and C. V. Deutsch, "Multivariate gaussian distribution", *Geostatistics*, vol. 470, 2022.

[85] S. J. Taylor and B. Letham, "Forecasting at scale", *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.

[86] B. W. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme", *Biochimica et Biophysica Acta (BBA)-Protein Structure*, vol. 405, no. 2, pp. 442–451, 1975.

[87] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy", *International journal of forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

[88] 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.

[89] N. Bellotto and H. Hu, "Computationally efficient solutions for tracking people with a mobile robot: An experimental evaluation of bayesian filters", *Autonomous Robots*, vol. 28, pp. 425–438, 2010.

[90] M. Oliveira, L. Torgo, and V. S. Costa, "Evaluation procedures for forecasting with spatio-temporal data", in *Joint European Conf. on Machine Learning and Knowledge Discovery in Databases*, 2018.

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

[92] R. Senanayake and F. Ramos, "Directional grid maps: Modeling multimodal angular uncertainty in dynamic environments", in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2018, pp. 3241–3248.

[93] S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[94] T. Krajník, J. P. Fentanes, M. Hanheide, and T. Duckett, "Persistent localization and life-long mapping in changing environments using the frequency map enhancement", in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2016, pp. 4558–4563.

[95] S. Molina, G. Cielniak, and T. Duckett, "Robotic exploration for learning human motion patterns", *IEEE Transactions on Robotics*, vol. 38, no. 2, pp. 1304–1318, 2021.

[96] L. Halodová, E. Dvořáková, F. Majer, *et al.*, "Adaptive image processing methods for outdoor autonomous vehicles", in *International Conference on Modelling and Simulation for Autonomous Systesm*, Springer, 2018, pp. 456–476.

[97] M. Rektoris, "Anomaly detection in periodic stochastic phenomena", B.S. thesis, České vysoké učení technické v Praze. Vypočetní a informační centrum., 2021.

[98] C. Coppola, T. Krajník, T. Duckett, and N. Bellotto, "Learning temporal context for activity recognition", in *22nd European Conference on Artificial Intelligence (ECAI)*, IOS Press, vol. 285, 2016, p. 107.

[99] J. Blaha, "Inferring temporal models of people presence from environment structrure", B.S. thesis, Czech Technical University in Prague, 2020.

[100] Laboratory of Chronorobotics, *Kdynakoupit*, `https://kdynakoupit.cz/`, Accessed: 2022-02-02, 2019.

[101] R. Triebel, K. Arras, R. Alami, *et al.*, "Spencer: A socially aware service robot for passenger guidance and help in busy airports", in *Field and service robotics*, Springer, 2016, pp. 607–622.

[102] S. Coşar, M. Fernandez-Carmona, R. Agrigoroaie, *et al.*, "Enrichme: Perception and interaction of an assistive robot for the elderly at home", *International Journal of Social Robotics*, vol. 12, no. 3, pp. 779–805, 2020.

[103] P. Neubert, N. Sünderhauf, and P. Protzel, "Appearance change prediction for long-term navigation across seasons", in *2013 European Conference on Mobile Robots*, IEEE, 2013, pp. 198–203.

[104] L. Calderita, A Vega, P Bustos, and P Núñez, "A new human-aware robot navigation framework based on time-dependent social interaction spaces: An application to assistive robots in caregiving centers", *Robotics and Autonomous Systems*, vol. 145, p. 103 873, 2021.

[105] M. Quigley, K. Conley, B. P. Gerkey, *et al.*, "ROS: An open-source robot operating system", in *ICRA Workshop on Open Source Software*, 2009.

[106] D. Hebesberger, T. Koertner, C. Gisinger, and J. Pripfl, "A long-term autonomous robot at a care hospital: A mixed methods study on social acceptance and experiences of staff and older adults", *International Journal of Social Robotics*, vol. 9, no. 3, pp. 417–429, 2017.

[107] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey", *The International Journal of Robotics Research*, vol. 39, no. 8, pp. 895–935, 2020.

[108] L. Palmieri, R. Andrey, J. Mainprice, *et al.*, "Guest editorial: Introduction to the special issue on long-term human motion prediction", *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5613–5617, 2021.

[109] T. Krajník, T. Vintr, G. Broughton, *et al.*, "Chronorobotics: Representing the structure of time for service robots", in *Proceedings of the 2020 4th International Symposium on Computer Science and Intelligent Control*, 2020, pp. 1–8.

[110] A. Ravankar, A. A. Ravankar, Y. Hoshino, M. Watanabe, and Y. Kobayashi, "Safe mobile robot navigation in human-centered environments using a heat map-based path planner", *Artificial Life and Robotics*, vol. 25, no. 2, pp. 264–272, 2020.

[111] T. Nishio and M. Niitsuma, "Environmental map building to describe walking dynamics for determination of spatial feature of walking activity", in *2019 IEEE 28th international symposium on industrial electronics (ISIE)*, IEEE, 2019, pp. 2315–2320.

[112] L. McCalman, S. O'Callaghan, and F. Ramos, "Multi-modal estimation with kernel embeddings for learning motion models", *2013 Ieee International Conference On Robotics And Automation (Icra)*, pp. 2845–2852, 2013.

[113] T. Kucner, M. Magnusson, E. Schaffernicht, V. Hernandez Bennetts, and A. Lilienthal, "Tell me about dynamics!: Mapping velocity fields from sparse samples with semi-wrapped gaussian mixture models", in *Robotics: Science and Systems Conference (RSS 2016), Workshop: Geometry and Beyond-Representations, Physics, and Scene Understanding for Robotics, University of Michigan, Ann Arbor, Ml, USA, June 18-22, 2016*, 2016.

[114] A. Roy, S. K. Parui, and U. Roy, "A mixture model of circular-linear distributions for color image segmentation", *International Journal of Computer Applications*, vol. 58, no. 9, 2012.

[115] C. S. Swaminathan, T. P. Kucner, M. Magnusson, L. Palmieri, and A. J. Lilienthal, "Down the cliff: Flow-aware trajectory planning under motion pattern uncertainty", in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2018, pp. 7403–7409.

[116] M. Kollmitz, K. Hsiao, J. Gaa, and W. Burgard, "Time dependent planning on a layered social cost map for human-aware robot navigation", in *2015 European Conference on Mobile Robots (ECMR)*, IEEE, 2015, pp. 1–6.

[117] R. Senanayake, M. Toyungyernsub, M. Wang, M. J. Kochenderfer, and M. Schwager, "Directional primitives for uncertainty-aware motion estimation in urban environments", in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2020, pp. 1–6.

[118] L. Nardi and C. Stachniss, "Long-term robot navigation in indoor environments estimating patterns in traversability changes", in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020, pp. 300–306.

[119] F. Kubiš, "Application of spatiotemporal modeling used in robotics for demand forecast", B.S. thesis, České vysoké učení technické v Praze. Vypočetní a informační centrum., 2020.

[120] J. Blaha, "Inferring temporal models of people presence from environment structrure", B.S. thesis, České vysoké učení technické v Praze. Vypočetní a informační centrum., 2020.

[121] H. Zhang and Z. Zheng, "Simulating nonstationary spatio-temporal poisson processes using the inversion method", in *2020 Winter Simulation Conference (WSC)*, IEEE, 2020, pp. 492–503.

[122] A. D. Cliff and J. K. Ord, "Model building and the analysis of spatial pattern in human geography", *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 37, no. 3, pp. 297–348, 1975. DOI: 111/bioe.12538.

[123] X. Shi and D.-Y. Yeung, "Machine learning for spatiotemporal sequence forecasting: A survey", *arXiv preprint arXiv:1808.06865*, 2018.

[124] W. Zhi, R. Senanayake, L. Ott, and F. Ramos, "Spatiotemporal learning of directional uncertainty in urban environments with kernel recurrent mixture density networks", *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4306–4313, 2019.

[125] R. Senanayake, K. B. Hatch, J. Zheng, and M. J. Kochenderfer, "3d radar velocity maps for uncertain dynamic environments", *arXiv preprint arXiv:2107.11039*, 2021.

[126] M. Leonard, "Směrové statistiky v predikci kvaziperiodickỳch časovỳch řad", B.S. thesis, České vysoké učení technické v Praze. Vypočetní a informační centrum., 2020.

[127] M. Stuede and M. Schappler, "Non-parametric modeling of spatio-temporal human activity based on mobile robot observations", *arXiv preprint arXiv:2203.06911*, 2022.

[128] D. Brščić, T. Kanda, T. Ikeda, and T. Miyashita, "Person tracking in large public spaces using 3-d range sensors", *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 6, pp. 522–534, 2013.

[129] S. T. O'Callaghan, S. P. N. Singh, A. Alempijevic, and F. T. Ramos, "Learning navigational maps by observing human motion patterns", *2011 Ieee International Conference On Robotics And Automation (Icra)*, 2011.

[130] E. O. Massey, *Comparative Analysis of Techniques for Spatio-Temporal World Modeling*. Hochschule Bonn-Rhein-Sieg, 2019.

[131] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures", *IEEE signal processing magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[132] Z. Talebpour, I. Navarro, and A. Martinoli, "On-board human-aware navigation for indoor resource-constrained robots: A case-study with the ranger", in *2015 IEEE/SICE International Symposium on System Integration (SII)*, IEEE, 2015, pp. 63–68.

[133] I. Kostavelis, A. Kargakos, D. Giakoumis, and D. Tzovaras, "Robot's workspace enhancement with dynamic human presence for socially-aware navigation", in *international conference on computer vision systems*, Springer, 2017, pp. 279–288.

[134] T. Dunning, *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press, 2012.

[135] B. Okal and K. O. Arras, "Learning socially normative robot navigation behaviors with bayesian inverse reinforcement learning", in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2016, pp. 2889–2895.

[136] D. Hebesberger, T. Koertner, C. Gisinger, J. Pripfl, and C. Dondrup, "Lessons learned from the deployment of a long-term autonomous robot as companion in physical therapy for older adults with dementia a mixed methods study", in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE, 2016, pp. 27–34.

[137] T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara, and K. Murase, "Development of human support robot as the research platform of a domestic mobile manipulator", *ROBOMECH Journal*, vol. 6, no. 4, 2019.

[138] I. Ko, B. Kim, and F. C. Park, "Randomized path planning on vector fields", *The International Journal of Robotics Research*, vol. 33, no. 13, pp. 1664–1682, 2014.

[139] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm", *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[140] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve.", *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[141] Y. Dodge, "The concise encyclopedia of statistics", in Springer Science & Business Media, 2008, ch. Chi-Square Distance.

[142] A. Bravais, *Analyse mathématique sur les probabilités des erreurs de situation d'un point*. Impr. Royale, 1844.

[143] L. Mudrova *et al.*, "An integrated control framework for long-term autonomy in mobile service robots", in *2015 European Conference on Mobile Robots (ECMR)*, IEEE, 2015, pp. 1–6.

[144] V. Vintrová, "Algorithms for finding local outliers", Ph.D. dissertation, Prague University of Economics and Business, 2021.

[145] C. S. Swaminathan, T. P. Kucner, M. Magnusson, *et al.*, "Benchmarking the utility of maps of dynamics for human-aware motion planning.", *Frontiers in Robotics and AI*, vol. 9, 2022.

[146] L. Castri, S. Mghames, M. Hanheide, and N. Bellotto, "Causal discovery of dynamic models for predicting human spatial interactions", in *Social Robotics: 14th International Conference, ICSR 2022, Florence, Italy, December 13–16, 2022, Proceedings, Part I*, Springer, 2023, pp. 154–164.

[147] A. Rudenko, T. P. Kucner, C. S. Swaminathan, R. T. Chadalavada, K. O. Arras, and A. J. Lilienthal, "Thör: Human-robot navigation data collection and accurate motion trajectories dataset", *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 676–682, 2020.

[148] T. Schreiter, T. R. de Almeida, Y. Zhu, *et al.*, "The magni human motion dataset: Accurate, complex, multi-modal, natural, semantically-rich and contextualized", *arXiv preprint arXiv:2208.14925*, 2022.

[149] T. Krajník, J. P. Fentanes, C. Dondrup, J. Santos, M. Hanheide, and T. Duckett. "Strands public datasets". Accessed: 2022-11-28. (2013), [Online]. Available: `https://lcas.lincoln.ac.uk/nextcloud/shared/datasets/index.html` (visited on 11/28/2022).

[150] T. Yang, Y. Li, C. Zhao, *et al.*, "3d tof lidar in mobile robotics: A review", *arXiv preprint arXiv:2202.11025*, 2022.

[151] G. Broughton, P. Linder, T. Rouček, T. Vintr, and T. Krajník, "Robust image alignment for outdoor teach-and-repeat navigation", in *2021 European Conference on Mobile Robots (ECMR)*, IEEE, 2021, pp. 1–6.

[152] T. Rouček, A. S. Amjadi, Z. Rozsypálek, *et al.*, "Self-supervised robust feature matching pipeline for teach and repeat navigation", *Sensors*, vol. 22, no. 8, 2022, ISSN: 1424-8220. DOI: 10.3390/s22082836. [Online]. Available: https://www.mdpi.com/1424-8220/22/8/2836.