2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 1-5, 2023. Detroit, USA

CLiFF-LHMP: Using Spatial Dynamics Patterns for Long-Term Human Motion Prediction

Yufei Zhu[1], Andrey Rudenko[2], Tomasz P. Kucner[3],
Luigi Palmieri[2], Kai O. Arras[2], Achim J. Lilienthal[1,4], Martin Magnusson[1]

2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 978-1-6654-9190-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IROS55552.2023.10342031

***Abstract*— Human motion prediction is important for mobile service robots and intelligent vehicles to operate safely and smoothly around people. The more accurate predictions are, particularly over extended periods of time, the better a system can, e.g., assess collision risks and plan ahead. In this paper, we propose to exploit *maps of dynamics* (MoDs, a class of general representations of place-dependent spatial motion patterns, learned from prior observations) for long-term human motion prediction (LHMP). We present a new MoD-informed human motion prediction approach, named CLiFF-LHMP, which is data efficient, explainable, and insensitive to errors from an upstream tracking system. Our approach uses CLiFF-map, a specific MoD trained with human motion data recorded in the same environment. We bias a constant velocity prediction with samples from the CLiFF-map to generate multi-modal trajectory predictions. In two public datasets we show that this algorithm outperforms the state of the art for predictions over very extended periods of time, achieving 45% more accurate prediction performance at 50s compared to the baseline.**

I. INTRODUCTION

Accounting for long-term human motion prediction (LHMP) is an important task for autonomous robots and vehicles to operate safely in populated environments [1]. Accurate prediction of future trajectories of surrounding people over longer periods of time is a key skill to improve motion planning, tracking, automated driving, human-robot interaction, and surveillance. Long-term predictions are useful to associate observed tracklets in sparse camera networks, or inform the robot of the long-term environment dynamics on the path to its goal [2, 3], for instance when following a group of people. Very long-term predictions are useful for global motion planning to produce socially-aware unobtrusive trajectories, and for coordinating connected multi-robot systems with sparse perception fields.

Human motion is complex and may be influenced by several hard-to-model factors, including social rules and norms, personal preferences, and subtle cues in the environment that are not represented in geometric maps. Accordingly, accurate motion prediction is very challenging [1]. Prediction on the very long-term scale (i.e., over 20 s into the future) is particularly hard as complex, large-scale environments influence human motion in a way that cannot be summarized and contained in the current state of the moving person or the observed interactions but rather have to be modelled explicitly [4].

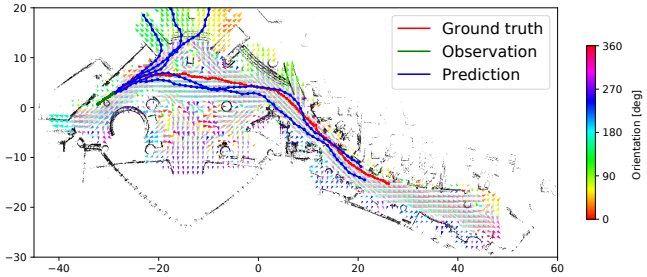

Fig. 1. Long-term (50 s) motion prediction result obtained with CLiFF-LHMP for one person in the ATC dataset. **Red** line: ground truth trajectory. **Green** line: observed trajectory. **Blue** lines: predicted trajectories. The CLiFF-map is shown with colored arrows.

In this paper, we examine and address the novel task of very long-term human motion prediction [5], aiming to predict human trajectories for up to 50 s into the future. Prior works have addressed human motion prediction using physics-, planning- and pattern-based approaches [1]. The majority of existing approaches, however, focuses on relatively short prediction horizons (up to 10 s) [6] and the popular ETH-UCY benchmark uses 4.8 s [1, 7, 8, 9].

To predict very long-term human motion, we exploit *maps of dynamics* (MoDs) that encode human dynamics as a feature of the environment. There are several MoD approaches for mapping velocities [10, 11, 12, 13, 14]. In this work, we use Circular Linear Flow Field map (CLiFF-map) [12], which captures multimodal statistical information about human flow patterns in a continuous probabilistic representation over velocities. The motion patterns represented in a CLiFF-map implicitly avoid collisions with static obstacles and follow the topological structure of the environment, e.g., capturing the dynamic flow through a hall into a corridor (see Fig. 1). In this paper we present a novel, MoD-informed prediction approach (CLiFF-LHMP)[1] that predicts stochastic trajectories by sampling from a CLiFF-map to guide a velocity filtering model [6]. Examples of prediction results are shown in Fig. 1.

In qualitative and quantitative experiments we demonstrate our CLiFF-LHMP approach is 45% more accurate than the baseline at 50 s, with average displacement error (ADE)

[1]AASS MRO lab, Örebro University, Sweden yufei.zhu@oru.se
[2]Bosch Corporate Research, Robert Bosch GmbH, Stuttgart, Germany andrey.rudenko@bosch.com
[3]Finnish Center for Artificial Intelligence, School of Electrical Engineering, Aalto University, Finland
[4]TU Munich, Germany

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017274 (DARKO).

[1]The approach is available at https://github.com/test-bai-cpu/CLiFF-LHMP

below $5\,\mathrm{m}$ up to $50\,\mathrm{s}$. In contrast to prior art in long-term environment-aware motion prediction [4], our method does not make any assumptions on the optimality of human motion and instead generalizes the features of human-space interactions from the learned MoD. Furthermore, our method does not require a list of goals in the environment as input, in contrast to prior planning-based prediction methods. Finally, our method can flexibly estimate the variable time end-points of human motion, predicting both short- and long-term trajectories, in contrast to the prior art which always predicts up to a fixed prediction horizon.

The paper is structured as follows: we review related work in Sec. II, describe the proposed approach in Sec. III, present our evaluation in Sec. IV, discuss the results in Sec. V and conclude in Sec. VI.

II. RELATED WORK

Human motion prediction has been studied extensively in recent years. With different prediction horizons, the human motion prediction problem can be divided into short-term (1–$2\,\mathrm{s}$), long-term (up to $20\,\mathrm{s}$) [1], and very long-term (which we define as over $20\,\mathrm{s}$). Several approaches address long-term motion prediction, e.g., full-body motion [5] or in the context of vehicle routing and GPS positioning [15, 16], but, to the best of our knowledge, very long-term prediction of dense navigation trajectories has not been addressed before.

One approach to predict long-term human motion is to account for various semantic attributes of the static environment. For instance, prior knowledge of potential goals in the environment can be used in planning-based methods. Ziebart et al. [17] and Karasev et al. [18] propose planning MDP-based approaches for long-term goal-directed global motion prediction. Rudenko et al. [4] extends this line of work by accounting for local social interactions, which is shown to outperform prior art in the long-term map-aware perspective.

Another popular approach to make long-term predictions is using clustering to represent observed long-term motion patterns, e.g., using expectation-maximization [19]. Chen et al. [20] use constrained gravitational clustering for dynamically grouping the observed trajectories, learning also how motion patterns change over time. Bera et al. [21] learn global and local motion patterns using Bayesian inference in real-time. One shortcoming of clustering-based methods is that they depend on complete trajectories as input. In many cases, e.g. in cluttered environments or from a first-person perspective [22], it is difficult to observe long trajectories, or cluster shorter tracklets and incomplete trajectories in a meaningful way.

Clustering-based methods directly model the distribution over full trajectories and are non-sequential. By contrast, transition-based approaches [23, 24, 25, 26, 27] describe human motion with causally conditional models and generate sequential predictions from learned local motion patterns.

Further, there are physics-based approaches that build a kinematic model without considering other forces that govern the motion. The constant velocity model (CVM) is a simple yet potent approach to predict human motion. Schöller et al. [28] have shown CVM to outperform several state-of-the-art neural predictors at the $4.8\,\mathrm{s}$ prediction horizon. On the other hand, CVM is not reliable for long-term prediction as it ignores all environment information.

Finally, many neural network approaches for motion prediction have been presented in recent years, based on LSTMs [29], GANs [30], CNNs [31], CVAEs [32] and transformers [33]. Most of these approaches focus on learning to predict stochastic interactions between diverse moving agents in the short-term perspective in scenarios where the effect of the environment topology and semantics is minimal. Our approach, on the other hand, targets specifically the long-term perspective, where the environment effects become critical for making accurate predictions.

Our approach to motion prediction leverages maps of dynamics (MoDs), which encode motion as a feature of the environment by building spatio-temporal models of the patterns followed by dynamic objects (such as humans) in the environment [14, 12]. There are several approaches for building maps of dynamics from observed motion. Some MoDs represent human dynamics in occupancy grid maps [24]. Another type of MoDs clusters human trajectories as mentioned above [19]. Chen et al. [34] present an approach that uses a dictionary learning algorithm to develop a part-based trajectory representation.

The above mentioned MoDs encode the direction but not the speed of motion. MoDs can also be based on mapping sparse velocity observations into flow models, which has the distinct advantage that the MoD can be built from incomplete or spatially sparse data. An example of this class of MoDs is the probabilistic Circular-Linear Flow Field map (CLiFF-map) [12] that we use in this paper. CLiFF-map uses a Gaussian mixture model (GMM) to describe multimodal flow patterns at each location. In this paper, we use sampled directions from the CLiFF-map to predict stochastic long-term human motion.

A method similar to ours is presented in Barata et al. [35]. It constructs a vector field that represents the most common direction at each point and predicts human trajectories by inferring the most probable sequence through this vector field. By contrast, our approach uses a probabilistic vector field that represents speed and direction jointly in a multimodal distribution. Further, the evaluation in Barata et al. [35] assumes a fixed prediction horizon of $4.8\,\mathrm{s}$, whereas we show our approach to estimate human motion more accurately than the state of the art for up to $50\,\mathrm{s}$.

III. METHOD

In this section, we first describe the CLiFF-map representation for site-specific motion patterns (Sec. III-A) and then present the CLiFF-LHMP approach for single-agent long-term motion prediction exploiting the information accumulated in a CLiFF-map (Sec. III-B).

A. Circular-Linear Flow Field Map (CLiFF-map)

To predict human trajectories we exploit the information about local flow patterns represented in a CLiFF-map as a

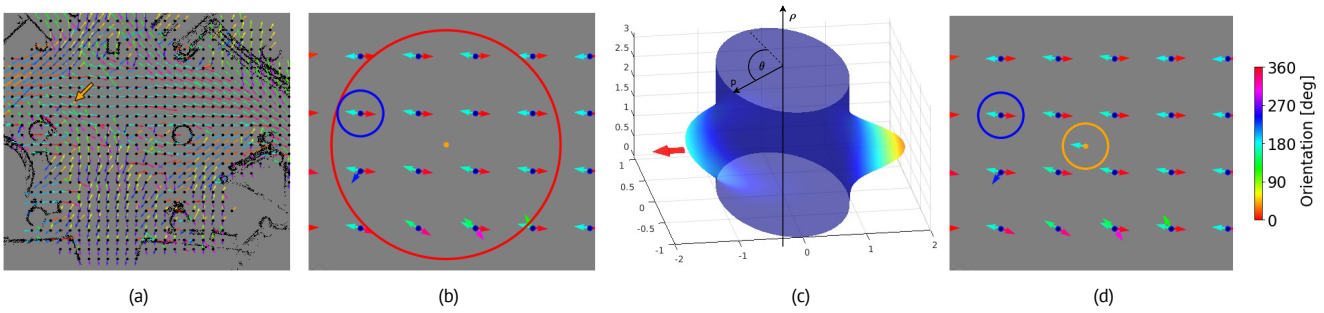

Fig. 2. Steps of sampling a direction θ_s from the CLiFF-map. **(a)** CLiFF-map built from the ATC data. The location to sample from is marked with an orange arrow. **(b)** Selection of SWGMMs in the CLiFF-map: The red circle contains all SWGMMs within r_s distance to the sampling location. From these SWGMMs, the SWGMM with the highest motion ratio is selected (marked with a blue circle). **(c)** The SWGMM distribution in the selected location wrapped on a unit cylinder. The speed is represented by the position along the ρ axis and the direction is θ. The probability is represented by the distance from the surface of the cylinder. A velocity vector (marked with a red arrow) is sampled from this SWGMM. **(d)** The direction value θ_s of the sampled velocity is shown in the sampled direction and marked with an orange circle.

multimodal, continuous distribution over velocities. CLiFF-map [12] is a probabilistic framework for mapping velocity observations (independently of their underlying physical processes), i.e., essentially a generalization of a vector field into a Gaussian mixture field. Each location in the map is associated with a Gaussian mixture model (GMM). A CLiFF-map represents motion patterns based on local observations and estimates the likelihood of motion at a given query location.

CLiFF-maps represent speed and direction jointly as velocity $\mathbf{V} = [\theta, \rho]^T$ using direction θ and speed ρ, where $\rho \in \mathbb{R}^+$, $\theta \in [0, 2\pi)$. As the direction θ is a circular variable and the speed is linear, a mixture of *semi-wrapped* normal distributions (SWNDs) is used in CLiFF-map. At a given location, the semi-wrapped probability density function (PDF) over velocities can be visualized as a function on a cylinder. Direction values θ are wrapped on the unit circle and the speed ρ runs along the length of the cylinder. An SWND $\mathcal{N}^{SW}_{\mathbf{\Sigma},\boldsymbol{\mu}}$ is formally defined as $\mathcal{N}^{SW}_{\mathbf{\Sigma},\boldsymbol{\mu}}(\mathbf{V}) = \sum_{k\in\mathbb{Z}} \mathcal{N}_{\mathbf{\Sigma},\boldsymbol{\mu}}([\theta,\rho]^T + 2\pi[k,0]^T)$, where $\mathbf{\Sigma}, \boldsymbol{\mu}$ denote the covariance matrix and mean value of the directional velocity $(\theta, \rho)^T$, and k is a winding number. Although $k \in \mathbb{Z}$, the PDF can be approximated adequately by taking $k \in \{-1, 0, 1\}$ for practical purposes [36]. To preserve the multi-modal characteristic of the flow, a semi-wrapped Gaussian mixture model (SWGMM) is used, which is a PDF represented as a weighted sum of J SWNDs: $p(\mathbf{V}|\xi) = \sum_{j=1}^{J} \pi_j \mathcal{N}^{SW}_{\mathbf{\Sigma_j},\boldsymbol{\mu_j}}(\mathbf{V})$, where $\boldsymbol{\xi} = \{\xi_j = (\boldsymbol{\mu}_j, \mathbf{\Sigma}_j, \pi_j) | j \in \mathbb{Z}^+\}$ denotes a finite set of components of the SWGMM, and π_j denotes the mixing factor and satisfies $0 \leq \pi_j \leq 1$.

B. Human Motion Prediction Using CLiFF-map

We frame the task of predicting a person's future trajectory as inferring a sequence of future states. The algorithm is presented in Alg. 1. With the input of an observation history of O_p past states of a person and a CLiFF-map Ξ, the algorithm predicts T_p future states. The length of the observation history is $O_s \in \mathbb{R}^+$ s, equivalent to $O_p > 0$ observation time steps. With the current time-step denoted as the integer $t_0 \geq 0$, the sequence of observed states is $\mathcal{H} = \langle s_{t_0-1}, ..., s_{t_0-O_p} \rangle$, where s_t is the state of a person at time-step t. A state is represented by 2D Cartesian coordinates (x, y), speed ρ and direction θ: $s = (x, y, \rho, \theta)$.

Algorithm 1: CLiFF-LHMP

Input: $\mathcal{H}$, x_{t_0}, y_{t_0}, Ξ
Output: $\mathcal{T}$

1 $\mathcal{T} = \{\}$
2 $\rho_{\text{obs}}, \theta_{\text{obs}} \leftarrow$ getObservedVelocity($\mathcal{H}$)
3 $s_{t_0} = (x_{t_0}, y_{t_0}, \rho_{\text{obs}}, \theta_{\text{obs}})$
4 **for** $t = t_0 + 1$, ..., $t_0 + T_p$ **do**
5 $\quad x_t, y_t \leftarrow$ getNewPosition(s_{t-1})
6 $\quad \theta_s \leftarrow$ sampleDirectionFromCLiFFmap(x_t, y_t, Ξ)
7 $\quad (\rho_t, \theta_t) \leftarrow$ predictVelocity(θ_s, ρ_{t-1}, θ_{t-1})
8 $\quad s_t \leftarrow (x_t, y_t, \rho_t, \theta_t)$
9 $\quad \mathcal{T} \leftarrow \mathcal{T} \cup s_t$
10 **return** $\mathcal{T}$

From the observed sequence $\mathcal{H}$, we derive the observed speed ρ_{obs} and direction θ_{obs} at time-step t_0 (line 2 of Alg. 1). Then the current state becomes $s_{t_0} = (x_{t_0}, y_{t_0}, \rho_{\text{obs}}, \theta_{\text{obs}})$ (line 3 of Alg. 1). The values of ρ_{obs} and θ_{obs} are calculated as a weighted sum of the finite differences in the observed states, as in the recent ATLAS benchmark [6]. With the same parameters as in [6], the sequence of observed velocities is weighted with a zero-mean Gaussian kernel with $\sigma = 1.5$ to put more weight on more recent observations, such that $\rho_{\text{obs}} = \sum_{t=1}^{O_p} v_{t_0-t} g(t)$ and $\theta_{\text{obs}} = \sum_{t=1}^{O_p} \theta_{t_0-t} g(t)$, where $g(t) = (\sigma\sqrt{2\pi} e^{\frac{1}{2}(\frac{t}{\sigma})^2})^{-1}$.

Given the current state s_{t_0}, we estimate a sequence of future states. Similar to past states, future states are predicted within a time horizon $T_s \in \mathbb{R}^+$ s. T_s is equivalent to $T_p > 0$ prediction time steps, assuming a constant time interval Δt between two predictions. Thus, the prediction horizon is $T_s = T_p \Delta t$. The predicted sequence is then denoted as $\mathcal{T} = \langle s_{t_0+1}, s_{t_0+2}, ..., s_{t_0+T_p} \rangle$.

To estimate $\mathcal{T}$, for each prediction time step, we sample a direction from the CLiFF-map at the current position (x_t, y_t) to bias the prediction with the learned motion patterns represented by the CLiFF-map. The main steps for each iteration are shown in lines 5–9 of Alg. 1.

For each iteration, we first compute the predicted position (x_t, y_t) at time step t from the state at the previous time step

Algorithm 2: sampleDirectionFromCLiFFmap(x, y, Ξ)

Input: x, y, Ξ
Output: θ_s
1 $\Xi_{\text{near}} \leftarrow$ getNearSWGMMs(x, y, Ξ)
2 $\xi \leftarrow$ selectSWGMM(Ξ_{near})
3 $\theta_s \leftarrow$ sampleDirectionFromSWGMM(ξ)
4 **return** θ_s

(line 5 of Alg. 1):

$$x_t = x_{t-1} + \rho_{t-1} \cos\theta_{t-1} \Delta t, \\ y_t = y_{t-1} + \rho_{t-1} \sin\theta_{t-1} \Delta t, \tag{1}$$

Afterwards, we estimate the new speed and direction using constant velocity prediction biased by the CLiFF-map. The bias impacts only the estimated direction of motion, speed is assumed to be unchanging.

To estimate direction at time t, we sample a direction from the CLiFF-map at location (x_t, y_t) in the function `sampleDirectionFromCLiFFmap()` (line 6 of Alg. 1). Alg. 2 outlines its implementation. The inputs of Alg. 2 are: the sample location (x, y) and the CLiFF-map Ξ of the environment. The sampling process is illustrated in Fig. 2. To sample a direction at location (x, y), from Ξ, we first get the SWGMMs Ξ_{near} whose distances to (x, y) are less than the sampling radius r_s (line 1 of Alg. 2). In a CLiFF-map, each SWGMM is associated with a motion ratio. To sample from the location with the highest intensity of human motions, in line 2, from Ξ_{near}, we select the SWGMM ξ with highest motion ratio. In line 3 of Alg. 2, from ξ, an SWND is sampled from the selected SWGMM, based on the mixing factor π. A velocity is drawn randomly from the sampled SWND. Finally, the direction of the sampled velocity is returned and used for motion prediction.

With the direction sampled from the CLiFF-map, we predict the velocity (ρ_t, θ_t) in line 7 of Alg. 1 assuming that a person tends to continue walking with the same speed as in the last time step, $\rho_t = \rho_{t-1}$, and bias the direction of motion with the sampled direction θ_s as:

$$\theta_t = \theta_{t-1} + (\theta_s - \theta_{t-1}) \cdot K(\theta_s - \theta_{t-1}), \tag{2}$$

where $K(\cdot)$ is a kernel function that defines the degree of impact of the CLiFF-map. We use a Gaussian kernel with a parameter β that represents the kernel width:

$$K(x) = e^{-\beta \|x\|^2}. \tag{3}$$

An example of velocity prediction results is shown in Fig. 3. With kernel K, we scale the CLiFF-map term by the difference between the direction sampled from the CLiFF-map and the current direction according to the CVM. The sampled direction is trusted less if it deviates more from the current direction. A larger value of β makes the proposed method behave more like a CVM, and with a smaller value of β, the prediction will follow the CLiFF-map more closely.

In the end of each iteration, we add s_t to the predicted trajectory $\mathcal{T}$ (line 9 of Alg. 1) and update t for the next iteration. After iterating for T_p times, the output is a sequence $\mathcal{T}$ of future states that represents the predicted trajectory.

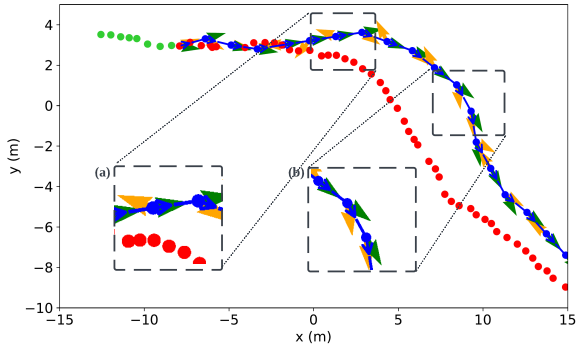

Fig. 3. Example predictions that visualize the adaptive influence of the CLiFF-map and the constant velocity model on the prediction, based on the sampled direction. **Green** dots show the observed past states $\mathcal{H}$, **red** dots show the ground truth future states and **blue** dots show the predicted states $\mathcal{T}$. In each predicted state, the **orange** arrow shows the sampled direction from the CLiFF-map θ_s and the **green** arrow shows the direction from the last time step θ_{t-1}. **Blue** arrows between predicted states show the direction of the predicted trajectory. In locations like (**a**) where the sampled CLiFF-map direction greatly opposes the CVM prediction, the CVM prediction is trusted more. In locations like (**b**) where the sampled CLiFF-map direction is close to the CVM prediction, the CVM prediction is biased more towards the CLiFF-map direction.

IV. EXPERIMENTS

This section describes the experimental setup for qualitative and quantitative evaluation of our CLiFF-LHMP approach. Accurate map-aware long-term motion predictions are typically addressed with Markov Decision Process (MDP) based methods [17, 18, 37, 38, 4]. Among them, as the baseline for CLiFF-LHMP, we chose the recent IS-MDP approach [4]. We also compare our method with the constant velocity predictor [28, 6].

We evaluate the predictive performance using the following two real-world datasets:

1) **THÖR** [39]: This dataset captures human motion in a room with static obstacles. It includes two settings: with one obstacle (denoted as THÖR1, see the top row in Fig. 9) and with three obstacles (denoted as THÖR3, see the bottom row in Fig. 9). The size of the room for data collection is 8.4×18.8 m.
2) **ATC** [40]: This dataset contains trajectories recorded in a shopping mall in Japan. The dataset covers a large indoor environment with total area of around $900\,\text{m}^2$. The map of the environment is shown in Fig. 1.

THÖR1 and THÖR3 both include four rounds of collected data. We use the first round to build the CLiFF-map and use the remaining three rounds for evaluation. After filtering out short trajectories (shorter than the observation horizon O_s) for evaluation, there are in total 247 trajectories in the THÖR1 dataset and 327 trajectories in the THÖR3 dataset. This gives us the train-to-test ratio of about 1 to 3 in both THÖR1 and THÖR3.

The ATC dataset consists of 92 days in total. For building the CLiFF-map, we used the data from the first day (Oct. 24th, 2012). From the remaining 91 days, again after filtering

Parameter	ATC	THÖR
observation horizon O_s	3.2 s	3.2 s
kernel parameter β	1	1
sampling radius r_s	1 m	0.5 m
prediction horizon T_s	1–50 s	0.4–12 s
prediction time step Δt	1 s	0.4 s
CLiFF-map resolution	1 m	0.5 m
kernal parameter σ	1.5	1.5
number of predicted trajectories k	20	20

TABLE I

PARAMETERS USED FOR EVALUATION IN THE ATC AND THÖR DATASETS

out trajectories shorter than the observation horizon O_s, we use 1 803 303 trajectories that have continuous motion.

We downsampled both datasets to $2.5\,\mathrm{Hz}$. For observation, we take $3.2\,\mathrm{s}$ (the first 8 positions) of the trajectory and use the remaining (up to $50\,\mathrm{s}$ or 125 positions) as the prediction ground truth. In the parameter analysis, we also evaluate the effect of setting the observation horizon to different values.

Given the area covered by the ATC dataset ($\sim 900\,\mathrm{m}^2$) and the THÖR dataset ($\sim 150\,\mathrm{m}^2$), the size and number of obstacles in THÖR dataset, and the trajectory lengths available in the datasets, we selected the parameters shown in Table I for our quantitative and qualitative experiments. Because the size of obstacles in the THÖR setting is less than $1\,\mathrm{m}$, we set the grid resolution to $0.5\,\mathrm{m}$ when building the CLiFF-map from the THÖR dataset, in contrast to $1\,\mathrm{m}$ in the ATC dataset. Also, we set the prediction time step Δt to $0.4\,\mathrm{s}$ for the cluttered THÖR dataset, in contrast to $1\,\mathrm{s}$ for the ATC dataset. In the parameter analysis we evaluate the impact of selecting Δt on prediction accuracy.

Sampling radius r_s and kernel β are the main parameters in CLiFF-LHMP. The value of r_s is set to a multiple of the CLiFF-map grid resolution. For biasing the current direction with the sampled one, we use the default value of β = 1 for both datasets. The impact of both parameters is evaluated in the experiments. Using the ATC dataset, we specifically evaluate the influence of the three parameters (see Fig. 6): observation horizon $O_s \in [1.2, 3.2]$ s, sampling radius $r_s \in [1, 3]$ m, and kernel parameter $\beta \in [0.5, 10]$. We also evaluated the influence of the prediction time step $\Delta t \in [0.4, 1.0]$ s using the THÖR dataset (see Fig. 7).

For the evaluation of the predictive performance we used the following metrics: *Average* and *Final Displacement Errors* (ADE and FDE) and *Top-k ADE/FDE*. ADE describes the error between points on the predicted trajectories and the ground truth at the same time step. FDE describes the error at the last prediction time step. *Top-k ADE/FDE* compute the displacements between the ground truth position and the closest of the k predicted trajectories. For each ground truth trajectory we predict k = 20 trajectories.

We stop prediction according to Alg. 1 when no dynamics data (i.e. SWGMMs) is available within the radius r_s from the sampled location (line 6). If one predicted trajectory stops before T_s, it will only be included in the ADE/FDE evaluation up to the last available predicted point. When predicting for each ground truth trajectory, the prediction horizon T_s is either equal to its length or $50\,\mathrm{s}$ for longer trajectories.

V. RESULTS

In this section, we present the results obtained in ATC and THÖR with our approach compared to two baselines. The performance evaluation is conducted using both quantitative and qualitative analysis, and we further investigate the approach's performance through a parameter analysis.

A. Quantitative Results

Figs. 4 and 5 show the quantitative results obtained in the ATC and THÖR datasets. We compare our CLiFF-LHMP approach with IS-MDP [4] and CVM. In the short-term perspective all approaches perform on par. The mean ADE is marginally lower for CVM compared to the other predictors below $6\,\mathrm{s}$ in ATC, below $10\,\mathrm{s}$ in THÖR1, and below $4\,\mathrm{s}$ in THÖR3. In THÖR3 there are more obstacles that people need to avoid, while THÖR1 and ATC include more open spaces. In open spaces without obstacles, a constant velocity prediction is often a very good short-term predictor [6]. For our approach which accounts for possible deviations from straight trajectories the ADE for short-term predictions is slightly higher. For prediction horizons less than $10\,\mathrm{s}$, IS-MDP performs better than CLiFF-LHMP. However, the IS-MDP method requires additional input (goal points and the obstacle map) and its performance strongly depends on both. In contrast, our approach makes predictions without explicit knowledge about goals and implicitly accounts for the obstacle layout, as well as the specific ways people navigate in the environment.

In long-term predictions above $10\,\mathrm{s}$, both CLiFF-LHMP and IS-MDP outperform the CVM method. Our approach is substantially better than IS-MDP when the prediction horizon is above $20\,\mathrm{s}$ since it implicitly exploits location-specific motion patterns, thus overcoming a known limitation of MDP-based methods [4]. Table II summarises the performance results of our method against the baseline approaches at the maximum prediction horizon. Our CLiFF-LHMP approach accurately predicts human motion up to $50\,\mathrm{s}$ with a mean ADE of $5\,\mathrm{m}$. At $50\,\mathrm{s}$ in the ATC dataset, our method achieves a 45% ADE and 55% FDE improvement in performance compared to IS-MDP. At $12\,\mathrm{s}$ in THÖR1 and THÖR3, our method achieves an improvement of 6.3% and 13.3% ADE (25.7%, 27.8% FDE) over IS-MDP, respectively.

Figs. 4 and 5 also show that the standard deviation of ADE and FDE is generally lower for CLiFF-LHMP predictions, compared to CVM and IS-MDP. This indicates that our approach makes more consistent predictions, both in the short- and long-term perspective.

B. Parameter Analysis

In the experiments with different observation horizons (see Fig. 6, left), our method performs robustly when the observation horizon is as low as $1.2\,\mathrm{s}$. In the experiments with different β values (see Fig. 6, middle), we find that $\beta = 1$ is a good trade-off. Lower β values make the

Dataset	Horizon	ADE / FDE (m)		
		CLiFF-LHMP	IS-MDP	CVM
ATC	50 s	**4.6 / 9.6**	8.4 / 21.3	12.4 / 27.1
THÖR1	12 s	**1.5 / 2.6**	1.6 / 3.5	1.8 / 3.8
THÖR3	12 s	**1.3 / 2.6**	1.5 / 3.6	2.8 / 6.1

TABLE II

LONG-TERM PREDICTION HORIZON RESULTS ON DIFFERENT DATASETS. WITH $O_s = 3.2\,\mathrm{s}$, ERROR REPORTED ARE ADE/FDE IN METERS.

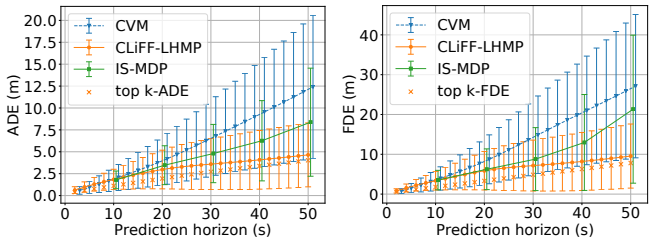

Fig. 4. ADE/FDE (mean ± one std. dev.) in the ATC dataset with prediction horizon 1–50 s.

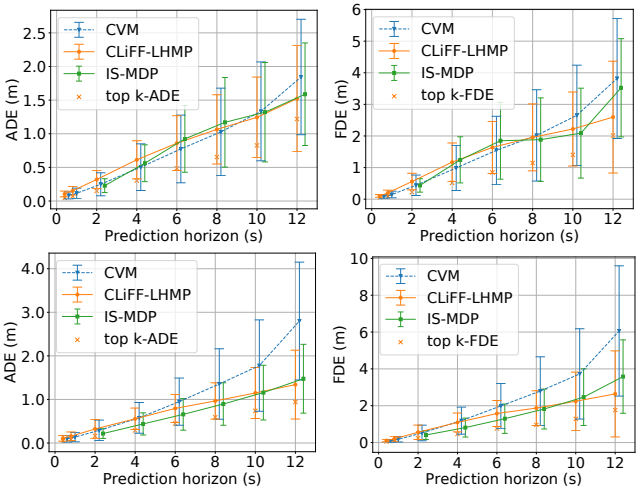

Fig. 5. ADE/FDE (mean ± one std. dev.) in the THÖR1 **(top)** and THÖR3 **(bottom)** dataset with prediction horizon 0.4–12 s.

predictor trust the CLiFF-map more, which can lead to jumps between distinct motion patterns. Setting β to a high value such as 10 slightly improves the performance in short-term predictions, however, as for the CVM model, the CLiFF-LHMP predictor with high values of β is prone to fail delivering long-term predictions. The reason is that we stop predicting when the CLiFF-map is not any longer available close to the predicted location. So, if more trust is put on the CVM component, many ground truth trajectories cannot be predicted successfully for long prediction times. When the planning horizon is set to 50 s, 84% of ground truth trajectories can be predicted successfully with $\beta = 1$, while with $\beta = 10$, the ratio drops to 52.3%. Also when the prediction is dominated by the CVM component, the top k-ADE/FDE scores are worse due to a reduced diversity of the predictions.

In the experiments with different values of the sampling radius r_s (see Fig. 6, right), we observed a stable prediction performance. Therefore, it is reasonable to set $r_s = 1$ in order to reduce the computation cost.

In our experiments with the prediction time step Δt, we observe robust performance with slight improvement when making higher frequency predictions (Δt =0.4 s vs. 1.0 s, see Fig. 7). Smaller Δt is recommended in cluttered environments, such as in the THÖR dataset. Making iterative predictions with a smaller time step naturally comes at the expense of computational cost increasing linearly for CLiFF-LHMP. Selecting a larger prediction time step Δt =1.0 s drops the performance in THÖR by only approx. 5% at the maximum prediction horizon, as compared to $\Delta t = 0.4\,\mathrm{s}$.

C. Qualitative Results

Figures 8 and 9 show qualitative results with example predictions. Our approach correctly captures the motion patterns in each scenario, utilizing the environment information during the prediction. Figure 9 shows that the predicted trajectories avoid the obstacles, even though an obstacle map is not used for predictions. Furthermore, using maps of dynamics built from the observations of human motion makes it possible to predict motion through regions which appear as obstacles in an occupancy map, for example across stairs and through narrow passages (see Fig. 8). Similarly, using the MoD input keeps predictions in more intensively used areas of the environment, avoiding semantically-insignificant and empty regions, e.g., corners of the room (see Fig. 9).

VI. CONCLUSIONS

In this paper we present the idea to use *Maps of Dynamics* (MoDs) for long-term human motion prediction. By using MoDs, motion prediction can utilize previously observed spatial motion patterns that encode important information about spatial motion patterns in a given environment. We present the CLiFF-LHMP approach to predict long-term motion using a CLiFF-map – a probabilistic representation of a velocity field from isolated and possibly sparse flow information (i.e. complete trajectories are not required as input). In our approach, we sample directional information from a CLiFF-map to bias a constant velocity prediction.

We evaluate CLiFF-LHMP with two publicly available real-world datasets, comparing it to several baseline approaches. The results demonstrate that our approach can predict human motion in complex environments over very long time horizons. Our approach performs on-par with the state of the art for shorter periods (10 s) and significantly outperforms it in terms of ADE and FDE for longer periods of up to 50 s. We also showed that our method makes more consistent predictions and is not strongly sensitive to the observation horizon. By exploiting the learned motion patterns encoded in the CLiFF MoD, our method can implicitly infer common goal points and correctly predict trajectories that follow the complex topology of the environment, e.g., navigating around corners or obstacles, or passing through narrow passages such as doors.

Future work will include experimenting with other types of MoDs and motion prediction methods, sampling speed in addition to direction from the MoD, extending CLiFF-LHMP to multi-agent prediction, extending the evaluation to

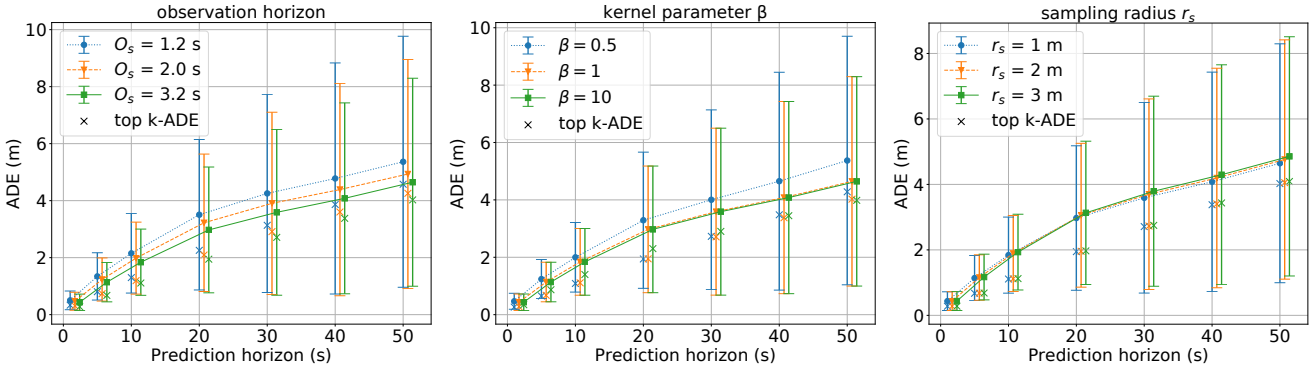

Fig. 6. Parameter analysis on the ATC dataset, showing the ADE (mean ± one std. dev.) over different prediction horizons vs the observation horizon O_s **(left)**, kernel parameter β **(middle)** and sampling radius r_s **(right)**.

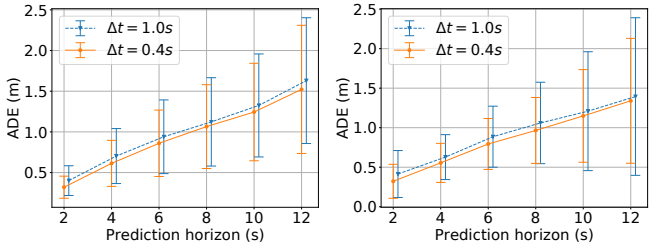

Fig. 7. Prediction time step Δt analysis on THÖR1 **(left)** and THÖR3 **(right)** datasets, showing the ADE (mean ± one std. dev.) over different prediction horizons.

outdoor datasets, as well as estimating confidence values for the predicted trajectories.

References

[1] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras. "Human motion trajectory prediction: A survey". In: *Int. J. of Robotics Research* 39.8 (2020), pp. 895–935.

[2] L. Palmieri, T. P. Kucner, M. Magnusson, A. J. Lilienthal, and K. O. Arras. "Kinodynamic motion planning on Gaussian mixture fields". In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. IEEE. 2017, pp. 6176–6181.

[3] 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: *Proc. of the IEEE Int. Conf. on Intell. Robots and Syst. (IROS)*. IEEE. 2018, pp. 7403–7409.

[4] A. Rudenko, L. Palmieri, A. J. Lilienthal, and K. O. Arras. "Human Motion Prediction under Social Grouping Constraints". In: *Proc. of the IEEE Int. Conf. on Intell. Robots and Syst. (IROS)*. 2018.

[5] J. Tanke, C. Zaveri, and J. Gall. "Intention-based Long-Term Human Motion Anticipation". In: *2021 International Conference on 3D Vision (3DV)*. IEEE. 2021, pp. 596–605.

[6] A. Rudenko, L. Palmieri, W. Huang, A. J. Lilienthal, and K. O. Arras. "The Atlas Benchmark: an Automated Evaluation Framework for Human Motion Prediction". In: *Proc. of the IEEE Int. Symp. on Robot and Human Interactive Comm. (RO-MAN)*. 2022.

[7] J. Amirian, J.-B. Hayet, and J. Pettré. "Social ways: Learning multi-modal distributions of pedestrian trajectories with GANs". In: *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR) Workshops*. 2019.

[8] B. Pang, T. Zhao, X. Xie, and Y. N. Wu. "Trajectory prediction with latent belief energy-based model". In: *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR)*. 2021, pp. 11814–11824.

[9] T. Gu, G. Chen, J. Li, C. Lin, Y. Rao, J. Zhou, and J. Lu. "Stochastic Trajectory Prediction via Motion Indeterminacy Diffusion". In: *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR)*. 2022, pp. 17113–17122.

[10] S. Molina, G. Cielniak, T. Krajník, and T. Duckett. "Modelling and Predicting Rhythmic Flow Patterns in Dynamic Environments". In: *Annual Conf. Towards Autonom. Rob. Syst.* Springer. 2018, pp. 135–146.

[11] T. Krajník, J. P. Fentanes, J. M. Santos, and T. Duckett. "FreMEn: Frequency Map Enhancement for Long-Term Mobile Robot Autonomy in Changing Environments". In: *IEEE Trans. on Robotics (TRO)* 33.4 (2017), pp. 964–977.

[12] T. P. Kucner, M. Magnusson, E. Schaffernicht, V. H. Bennetts, and A. J. Lilienthal. "Enabling Flow Awareness for Mobile Robots in Partially Observable Environments". In: *IEEE Robotics and Automation Letters* 2.2 (2017), pp. 1093–1100.

[13] W. Zhi, R. Senanayake, L. Ott, and F. Ramos. "Spatiotemporal Learning of Directional Uncertainty in Urban Environments With Kernel Recurrent Mixture Density Networks". In: *IEEE Robotics and Automation Letters* 4.4 (2019), pp. 4306–4313.

[14] T. P. Kucner, A. J. Lilienthal, M. Magnusson, L. Palmieri, and C. S. Swaminathan. *Probabilistic mapping of spatial motion patterns for mobile robots*. Springer, 2020.

[15] P.-C. Cheng, K. C. Lee, M. Gerla, and J. Härri. "GeoDTN+ Nav: geographic DTN routing with navigator prediction for urban vehicular environments". In: *Mobile Networks and Applications* 15.1 (2010), pp. 61–82.

[16] Z. Xiao, P. Li, V. Havyarimana, G. M. Hassana, D. Wang, and K. Li. "GOI: A novel design for vehicle positioning and trajectory prediction under urban environments". In: *IEEE Sensors Journal* 18.13 (2018), pp. 5586–5594.

[17] B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, and S. Srinivasa. "Planning-based prediction for pedestrians". In: *Proc. of the IEEE Int. Conf. on Intell. Robots and Syst. (IROS)*. 2009, pp. 3931–3936.

[18] V. Karasev, A. Ayvaci, B. Heisele, and S. Soatto. "Intent-aware long-term prediction of pedestrian motion". In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2016, pp. 2543–2549.

[19] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun. "Learning motion patterns of people for compliant robot motion". In: *Int. J. of Robotics Research* 24.1 (2005), pp. 31–48.

[20] Z. Chen, D. C. K. Ngai, and N. H. C. Yung. "Pedestrian behavior prediction based on motion patterns for vehicle-to-pedestrian collision avoidance". In: *Proc. of the IEEE Int. Conf. on Intell. Transp. Syst. (ITSC)*. 2008, pp. 316–321.

[21] A. Bera, S. Kim, T. Randhavane, S. Pratapa, and D. Manocha. "GLMP-realtime pedestrian path prediction using global and local movement patterns". In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2016, pp. 5528–5535.

[22] C. Dondrup, N. Bellotto, F. Jovan, and M. Hanheide. "Real-Time Multisensor People Tracking for Human-Robot Spatial Interaction". In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA), Workshop on ML for Social Robo*. IEEE. 2015.

[23] S. Thompson, T. Horiuchi, and S. Kagami. "A probabilistic model of human motion and navigation intent for mobile robot path planning". In: *Proc. of the IEEE Int. Conf. on Autonomous Robots and Agents (ICARA)*. 2009, pp. 663–668.

[24] Z. Wang, P. Jensfelt, and J. Folkesson. "Modeling spatial-temporal dynamics of human movements for predicting future trajectories". In: *Workshop Proc. of the AAAI Conf. on Artificial Intelligence*

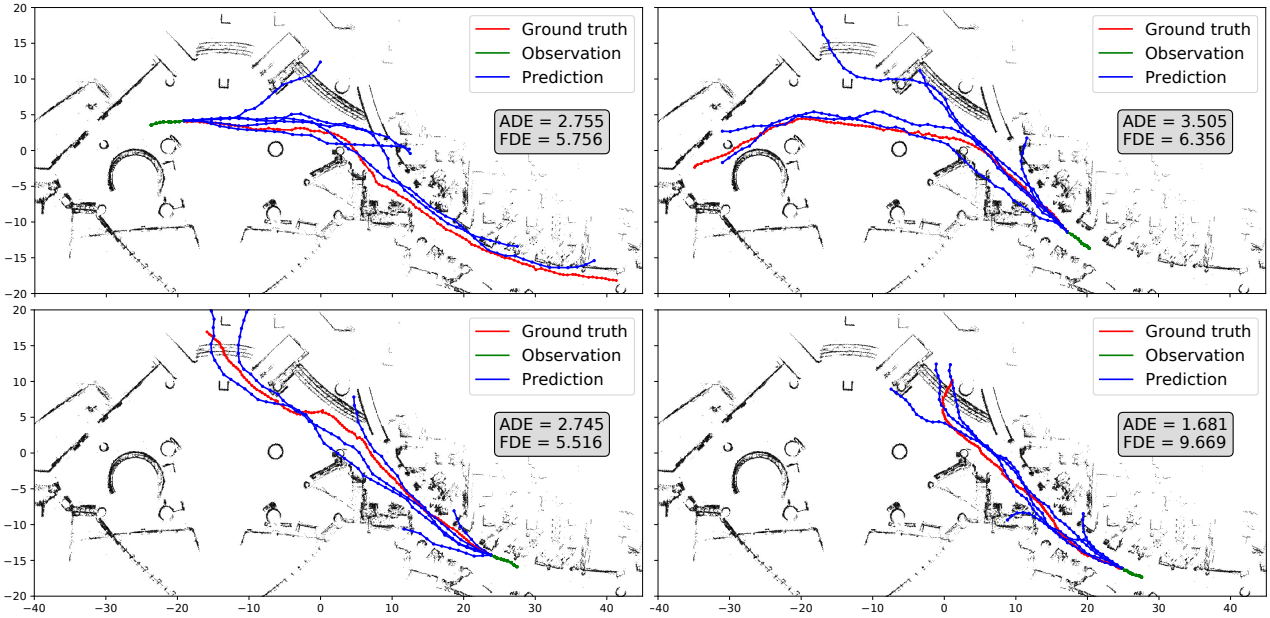

Fig. 8. Predictions in ATC with $T_s = 50$ s. **Red** line shows the ground truth trajectory. **Green** line shows the observed trajectory and **blue** lines show the predicted trajectories. Note that we correctly predict trajectories crossing obstacles such as stairs (top of the map) and exits (left of the map).

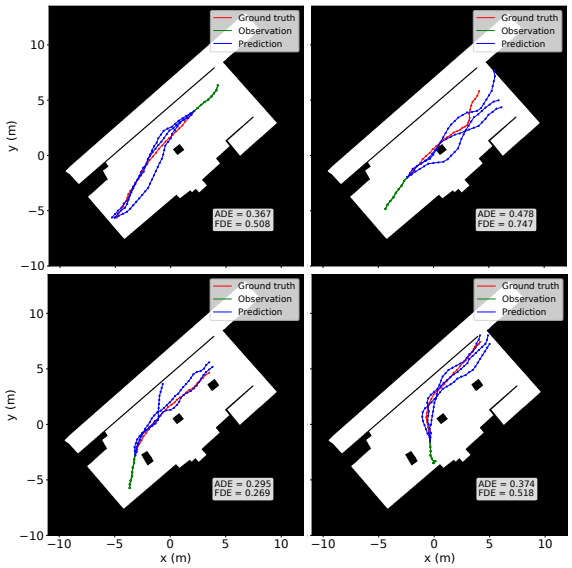

Fig. 9. Predictions in THÖR1 **(top)** and THÖR3 **(bottom)** with $T_s = 12$ s. **Red** line shows the ground truth trajectory. **Green** line shows the observed trajectory and **blue** lines show the predicted future trajectories

"Knowledge, Skill, and Behavior Transfer in Autonomous Robots". 2015.

[25] L. Ballan, F. Castaldo, A. Alahi, F. Palmieri, and S. Savarese. "Knowledge transfer for scene-specific motion prediction". In: *Proc. of the Europ. Conf. on Comp. Vision (ECCV)*. Springer. 2016, pp. 697–713.

[26] T. P. Kucner, J. Saarinen, M. Magnusson, and A. J. Lilienthal. "Conditional transition maps: Learning motion patterns in dynamic environments". In: *Proc. of the IEEE Int. Conf. on Intell. Robots and Syst. (IROS)*. 2013, pp. 1196–1201.

[27] J. Saarinen, H. Andreasson, and A. J. Lilienthal. "Independent Markov chain occupancy grid maps for representation of dynamic environment". In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2012, pp. 3489–3495.

[28] C. Schöller, V. Aravantinos, F. Lay, and A. Knoll. "What the constant velocity model can teach us about pedestrian motion prediction". In: *IEEE Robotics and Automation Letters* 5.2 (2020), pp. 1696–1703.

[29] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. "Social LSTM: Human trajectory prediction in crowded spaces". In: *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR)*. 2016, pp. 961–971.

[30] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, and S. Savarese. "SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints". In: *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR)*. 2019, pp. 1349–1358.

[31] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel. "Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction". In: *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR)*. 2020, pp. 14424–14432.

[32] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone. "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data". In: *European Conference on Computer Vision*. Springer. 2020, pp. 683–700.

[33] F. Giuliari, I. Hasan, M. Cristani, and F. Galasso. "Transformer networks for trajectory forecasting". In: *Proc. of the IEEE Int. Conf. on Pattern Recognition*. IEEE. 2021, pp. 10335–10342.

[34] Y. F. Chen, M. Liu, and J. P. How. "Augmented dictionary learning for motion prediction". In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2016, pp. 2527–2534.

[35] C. Barata, J. C. Nascimento, J. M. Lemos, and J. S. Marques. "Sparse motion fields for trajectory prediction". In: *Pattern Recognition* 110 (2021), p. 107631.

[36] K. V. Mardia and P. E. Jupp. *Directional Statistics*. Wiley, 2008.

[37] E. Rehder, F. Wirth, M. Lauer, and C. Stiller. "Pedestrian prediction by planning using deep neural networks". In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2018, pp. 1–5.

[38] A. Rudenko, L. Palmieri, and K. O. Arras. "Joint Prediction of Human Motion Using a Planning-Based Social Force Approach". In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2018, pp. 1–7.

[39] 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". In: *IEEE Robotics and Automation Letters* 5.2 (2020), pp. 676–682.

[40] D. Brščić, T. Kanda, T. Ikeda, and T. Miyashita. "Person tracking in large public spaces using 3-D range sensors". In: *IEEE Trans. on Human-Machine Systems* 43.6 (2013), pp. 522–534.