Multi-scale Conditional Transition Map: Modeling Spatial-temporal Dynamics of Human Movements with Local and Long-term Correlations

Zhan Wang, Patric Jensfelt and John Folkesson

Abstract—This paper presents a novel approach to modeling the dynamics of human movements with a grid-based representation. The model we propose, termed as Multi-scale Conditional Transition Map (MCTMap), is an inhomogeneous HMM process that describes transitions of human location state in spatial and temporal space. Unlike existing work, our method is able to capture both local correlations and long-term dependencies on faraway initiating events. This enables the learned model to incorporate more information and to generate an informative representation of human existence probabilities across the grid map and along the temporal axis for intelligent interaction of the robot, such as avoiding or meeting the human. Our model consists of two levels. For each grid cell, we formulate the local dynamics using a variant of the left-to-right HMM, and thus explicitly model the exiting direction from the current cell. The dependency of this process on the entry direction is captured by employing the Input-Output HMM (IOHMM). On the higher level, we introduce the place where the whole trajectory originated into the IOHMM framework forming a hierarchical input structure to capture long-term dependencies. The capabilities of our method are verified by experimental results from 10 hours of data collected in an office corridor environment.

I. INTRODUCTION

For robots engaged in operations in real-world scenarios with dynamic objects, such as people, one of the major challenges is to learn a model of the long-term behaviors of these moving objects. With a proper model summarizing their complex activity patterns in time and space, the robot is able to make local predictions and reason about future movements based on limited observations to improve its interaction with the humans or other moving objects, for example by staying out of their intended path or meeting people where they regularly appear.

In this work, we tackle this problem by proposing the Multi-scale Conditional Transition Map (MCTMap), as a global transition model summarizing human activity events based on a grid-based representation. The occupancy grid map [1] and its variants have been employed in modeling dynamics in the environment by many methods, for example [2], [3], [4], [5], [6], [7]. In the work [3], the occupancy grid map is generalized and the static environment assumption is relaxed by introducing the Hidden Markov Model (HMM). The developed algorithm exhibits excellent capability to distinguish between high dynamic (such as people), semi-dynamic (such as chairs) and static objects, showing great promise of using the grid representation for dynamic environments. However, in most methods, each cell is modeled independently and the correlations across cells are not considered. For example in [4] and [5], the activity in each individual cell are modeled as either an independent two-state Markov chain or an inhomogeneous Poisson process.

We make the observation that the immediate moving tendency of people in human-living environments is dependent on both the manner of entering the current position and the faraway initiating event (starting moving from the place they stay), and we show that MCTMap is able to capture both of these correlations using a probabilistic framework. On the local level, the occurrence of entering events in a neighboring region contains valuable information about possible future events in the region of interest. This information was exploited in [6], but by way of the occupancy of cells rather than explicitly modeling the tendency of moving as in MCTMap, and therefore no useful object dynamics representations are shown to be generated for path planning. The Conditional Transition Maps (CTMaps) [8] directly models the local moving tendency but the value of conditional transition parameters are learned by counting entry and exit events, and no inference of future occurrence probabilities of observed objects are provided. On the higher level, people typically move under environmental constraints. For example in an office corridor environment, typical human movements start and end in functional places, termed as ‘resting places’ [9], such as offices, kitchens and bathrooms. Human movements on this topological level exhibits routine patterns. Obviously the events in high-level human behaviors have significant influence on the local movements in a specific location. In this work, we choose to incorporate the resting place from where the whole trajectory is started, which we term as the ‘starting place’. Neither of the above methods is able to handle the useful high-level information of this kind and we will show that the long-term dependency of the local statistical process on remote events can be captured due to the structured human behaviors.

Our method works by learning cell transitions dependent on both the adjacent grid cell that was occupied just before the currently occupied cell (i.e. the local direction) and the high-level starting place of the trajectory (i.e. the global source). The ‘Multi-scale’ in the name refers to the two level of sources that the local transition is conditioned on. On the
local level, for each cell we employ a variant of the left-to-right HMM to directly model the moving tendency to a neighbor cell. The correlation between exiting and entering the current cell is formulated by utilizing the IOHMM [10], [11]. On the global level, we introduce faraway initiating events, such as the resting place where the trajectory starts, into the hierarchical input of the IOHMM, and thus provide high-level guidance to the local process. In this paper, we present how learning and inference in the MCTMap are performed.

II. RELATED WORK

In the study of dynamics in the environment, many methods are based on an underlying static representation, but extending it into a timescale framework or applying recency-weighted techniques. Arbuckle et al. [2] extended the occupancy grid map by maintaining multiple occupancy representations corresponding to various timescales. With the resulting temporal occupancy grid (TOG), the dynamics of each cell is classified by the occupancy values over different timescales. Saarinen et al. [7] proposed a 3D modeling approach combining the normal distributions transform (NDT) and the occupancy grid map. The method has the ability of adapting to a dynamic environment with a recency-weighted strategy. Impressive results are provided for long-term applications in a large scale dynamic environment in a milk production plant.

The dynamic occupancy grid [3], which utilizes a HMM with a two-state Markov chain to model the occupancy of a cell, successfully released the assumption of static environments for the traditional occupancy grid maps. The dynamics of each cell is explicitly represented by the transition probabilities. While the method is inherently homogeneously averaging dynamics in a certain timescale, the online training procedure provides adaptive capabilities from the recent tendency. In [5], each cell is modeled as an independent two-state Markov chain, and the transition probabilities are modeled as two Poisson processes and learned in an online manner, approximated by the frequency of 'exit' and 'enter' events. The spatial affordance map [4] represents human activity events as a rate function of an inhomogenous spatial Poisson process. It provides a global long-term summary of human activity events and the model can be used as environment constraints for people tracking and other interaction with people. Different from [4], the MCTMap that we propose is a global transition model that can generate informative heat map like representations of occurrence probabilities for the entire space based on a few online observations.

Common to the above methods based on occupancy grids is that each individual cell is modeled independently. More closer to our proposed approach, in [6], the local transition probability is conditioned on events in the neighbor cell, however the moving tendency is not directly modeled and therefore no useful object dynamics representation can be generated for path planning. In the Conditional Transition Maps (CTMap) developed by Kucner et al. [8], the cross-cell spatial relation is modeled as a probability distribution of an object leaving to a neighbor cell conditioned on the entry direction. Cross-correlation is used to find entry and exit events and the value of conditional transition parameters are learned by counting these events, but no effective inference of object occurrence is provided. In contrast to our approach, none of these two methods model long-term dependencies, which contain significant amount of information in structured human-living environments.

Another line of research on modeling motion patterns is based on studying the trajectories. Morris et al. [12] provide a survey of techniques for trajectory-based activity analysis using vision in a wide range of environments. The work [13] addresses the problem of vision-based future action inference by using a semantic scene understanding technique and optimal control theory. Ikeda et al. [14] propose a method to predict the long-term behavior of pedestrians by estimating the final goal using the concept of sub-goal. The locations of sub-goals and the transitions among them are learned and the future positions are predicted based on initial observations. The sub-goal is similar to the resting place in our work, but they only model the human behavior on the higher level, ignoring motion details between sub-goals. Bennewitz et al. [9] characterize typical motion patterns in human moving trajectories in indoor environments by using the HMM. The probabilistic belief about the potential trajectories of persons are incorporated into the path planning process of a mobile robot by using the A* algorithm. Unlike our method, the belief of human future location is constrained on the trajectories that are modeled by the HMM. In our proposed method, we focus on the grid-based methods as our applications are based on the grid map representation.

III. MULTI-SCALE CONDITIONAL TRANSITION MAP

A. Modeling Cell-level Correlation Using IOHMM with Left-to-right Structure

Similar to the occupancy grid map representation, we divide the overall map into grid cells in the two dimensional space. Accordingly, the original data sequence is discretized in time domain such that the observed human moves only to a neighbor cell in one time step. The dynamics of the human movement is studied on the cell level.

![Fig. 1. An example of one-step movement in a corridor environment, moving out of the current cell C to R, with the entering cell being L, as indicated by the dashed arrowed line. The solid arrowed lines indicate possible moving directions from the current cell C.](image)

The continuous movement of an object can be considered as the combination of a sequence of one-step movements, which are defined as moving from the current cell to one of
the neighbor cells within one or more time steps subjective to temporal and spatial discretization. We make the observation that where to go in the next time step is highly related to the manner that the human entered the current cell. This indicates the existence of correlations in the motion across three cells, for example the motion in the order of L→C→R, as indicated by the dashed arrowed line in Figure 1.

We model the tendency of the object moving from the current cell to a neighbor cell by a non-ergodic HMM, and the structure facilitates capturing the spatial-temporal correlation between the two cells in a movement. The dependency on how the object entered the current cell is formulated by using the Input-Output HMM and introducing the entering direction as the input, covering the correlations across the three involved cells. Details of these two aspects of our model will be described in the following two subsections.

1) Left-to-right HMM: Based on the above analysis, for each cell, we model the one-step movement by a variant of the left-to-right HMM with five states. Different from normal left-to-right HMMs, our model has four absorbing states, instead of just one. Abusing the term ‘left-to-right’, we refer to the structure of our model as left-to-right in the rest of this paper for conciseness.

A small grid map shown in Figure 1 is attached to each cell in which the current cell is denoted as C. The process for an individual cell starts from the object entering the current cell. In each of the following time steps, the object can either stay in the current cell or move in one of the four directions (left, right, up, down) as shown in Figure 1, and the process ends when the object reaches any of the neighbor cells. The latent variable $x_t$ represents the location of the object in the five cells involved in the current one-step movement at time step $t$, and it can take five states $\{C, L, R, U, D\}$, corresponding to the five cells (center, left, right, up, down). Let $z_t$ be a random variable that represents the observation of the object in a cell at time step $t$, and it can take five states $\{C, L, R, U, D\}$. Due to the relatively large grid size we applied (0.3m in the experiments) and the temporal and spatial discretization in generating training sequences, the four directions are sufficient to describe the one-step movement of a human. When the grid size is small, the chances of moving directly into the diagonal cells increase and the states corresponding to the cells in the diagonal direction need to be added.

![Fig. 2. Transition diagram of our model of one-step movement by a variant of the left-to-right HMM.](image)

The process under consideration is only one move from the current cell, and thus the states $\{L, R, U, D\}$ are specified as ‘absorbing states’. Let $A_{lk}$ represent the state transition probability from the $l^{th}$ state to the $k^{th}$ state, where $l, k = 1, \ldots, 5$. Figure 2 shows the transition diagram.

The corresponding 5 by 5 transition matrix is sparse in that $A_{lk} = 0, (l > k)$ and $A_{lk} = 0, (l < k, l > 1)$. The ones on the main diagonal, $A_{lk} = 1, (l = k, l > 1)$, indicate that the corresponding states $\{L, R, U, D\}$ are absorbing states.

2) Input-Output HMM: As shown in Figure 1, the direction from which the object entered the current cell has a strong influence on the manner in which it moves out. Hence the transition probabilities in our model are also dependent on the location from which it entered the current cell and take the form of

$$p(x_{t+1}|x_t, U_{t+1}),$$

in which $U_{t+1}$ represents the observation of the location of the object in neighbor cells surrounding the current cell at time step $t – 1$. Figure 3 shows our formulation of the process using the Input-Output HMM in which the transition probability is conditioned on the input $U_{t+1}$. For the IOHMM process of the current cell the input $U_{t+1}$ is from the observations of the IOHMMs in neighbor cells. Close examination shows that the IOHMMs in different cells are independent conditioned on the observation set.

![Fig. 3. Structure of our model of one-step movement using the IOHMM.](image)

By using the IOHMM, the local spatial-temporal correlations across the three cells in the one-step movement are captured. More correlations can be captured by adding more observations of the location of the object further away from the current cell at earlier time steps into $U_{t+1}$, but at the cost of increased complexity. When the object stays in the current cell for more than one time step, the elements in $U_{t+1}$ become all zero. The correlation with the entry direction is still maintained due to the bias items of equation (4) in the next subsection. The correlation is only blurred when objects staying in the current cell come from different directions. However this is addressed by introducing the starting place into the input as described in Section III-B.

The latent state, $x_t$, depends on the input by way of the transition model. We believe that the transition probabilities are highly related to the manner by which the object enters the current cell, and thus can be used to capture the motion pattern of objects in the environment. As such, in our IOHMM based model, the transition probabilities are dependent on the observations of the object’s location in the neighbor cells.

In the process of the one-step movement that is centered in the current cell, at time step $t$, we connect $U_{t+1}$, which
represents the location observation at time step \( t - 1 \), and \( z_{t+1} \) in the form of an input-output pair in the IOHMM. However, this involves no efforts of tracking any object. The model is purely based on the observation of events in the neighborhood of the current cell at successive time steps.

It is worth noting that alternatively the above process can be formulated by using a second order HMM. However, in that case, the model will lose the clear structure of mapping input to transition probabilities in the IOHMM, which will be described in the next subsection.

3) Mapping Input to Transition Probabilities Using Neural Network: Each transition probability is a conditional distribution, through which the probability distribution of \( x_{t+1} \) is dependent on \( x_t \) and the input \( U_{t+1} \) in the form of

\[
A_{t+1,k} = p(x_{t+1,k}|x_t,U_{t+1}), \tag{2}
\]

where \( x_{t+1,k} \) denotes the latent variable at time step \( t+1 \) being in the \( k^{th} \) state.

Due to the relatively large grid size (0.3m as used in the experiments), the input \( U_{t+1} \) represents the observation of the object location in four neighbor cells (left, right, up, down) at time step \( t - 1 \). The observation is coded by four binary variables, \( z_{n1}^{t}, z_{n2}^{t}, z_{n3}^{t}, z_{n4}^{t} \). The input at time step \( t + 1 \) is

\[
U_{t+1} = [z_{n1}^{t+1}, z_{n2}^{t+1}, z_{n3}^{t+1}, z_{n4}^{t+1}]. \tag{3}
\]

where \( z_{n1}^{t+1} \) represents the observation of object in the \( i^{th} \) neighbor cell at time step \( t - 1 \). When the grid size is small then the input should include all eight neighbor cells.

The transition probabilities are formulated using a two layer neural network, which is defined following [15]. For clarity purpose, the time subscript is omitted in the rest of this subsection and the input \( U_{t+1} \) denoted as \( U_{t+1} = [u_{1}^{t+1}, u_{2}^{t+1}, u_{3}^{t+1}, u_{4}^{t+1}] \).

First, four linear combinations of the inputs are constructed in the form

\[
a_j = \sum_{i=1}^{4} w_{ji}^{(1)} u^{i} + w_{j0}^{(1)}, \tag{4}
\]

where \( j = 1, \ldots, 4 \), and the superscript \( (1) \) indicates the first layer of the network. The parameters \( w_{ji}^{(1)} \) and \( w_{j0}^{(1)} \) are referred to as weights and biases respectively.

Then each of \( a_j \), known as activation, is transformed by using a sigmoidal function

\[
b_j = h(a_j), \tag{5}
\]

to obtain the hidden units, \( b_j \), where \( j = 1, \ldots, 4 \).

Lastly the hidden units are linearly combined to obtain the transition probabilities

\[
A_{1k} = \sum_{j=1}^{4} w_{kj}^{(2)} b_j + w_{k0}^{(2)}, \tag{6}
\]

where \( k = 1, \ldots, 5 \). The weights and biases are collectively denoted as \( W = [w_{ji}^{(1)}, w_{j0}^{(1)}, w_{kj}^{(2)}, w_{k0}^{(2)}] \), where \( i, j = 1, \ldots, 4 \), \( k = 1, \ldots, 5 \).

4) Training Procedure: As the IOHMMs in different cells are independent conditioned on the observation set, the training for an individual IOHMM can be performed separately. We adapt the scheme in [10] with the generalized EM algorithm [16] to the special needs in training our model. Due to the special left-to-right structure of the HMM that we exploit, a single long-observation sequence is not suitable for training, because once any of the absorbing states is reached, the rest of the sequence provides no further information. We generate training data for each cell in the form of a set of short-observation sequences each starting from the current cell and ending in one of its neighbor cells. The procedure for training general HMMs is also modified as follows to handle the left-to-right structure [17].

- As shown by the transition diagram in Figure 2, the process always starts from the first state, \( C \), termed as the starting state, and hence the prior is set as \( \pi = (1, 0, 0, 0) \) and not reestimated.
- At the beginning of the forward-backward scheme [18], set the transition probabilities \( A_{lk} = 0, (l < k) \) and \( A_{lk} = 0, (l > k) \).
- The probability \( \beta(x_t) \) in the forward-backward scheme is defined as

\[
\beta(x_t) = p(z_{t+1}, \ldots, z_{T}|x_t, U_{t+1}, \ldots, U_T), \tag{7}
\]

where \( T \) is the total number of time steps in the current training sequence [10]. The initial condition \( \beta(x_T) \) is set as

\[
\beta(x_T) = \begin{cases} 1 & \text{if } x_T \in \{L, R, U, D\} \\ 0 & \text{if } x_T \in \{C\}. \end{cases}
\]

Suppose the training data for the current cell are a set of \( P \) short-observation sequences, \( \{(z_{1p}^T, U_{1p}^T(p)); p = 1, \ldots, P\} \). Each sequence consists of all observations

\[
z_{1p}(t) = \{(z_{1p}(0), \ldots, z_{1p}(T))\}
\]

and inputs \( U_{1p}^T(p) = \{U_1(p), \ldots, U_T(p)\} \) of the corresponding one-step movement. In the following the parameter \( p \) is omitted.

For each short-observation sequence, let \( \gamma(x_t) \) denote the marginal posterior distribution of the latent variable \( x_t \), and \( \xi(x_{t-1}, x_t) \) denote the joint posterior distribution of two successive latent variables, given the observations, the inputs, and the model parameters,

\[
\gamma(x_t) = p(x_t|z_{1t}^T, U_{1t}^T, \hat{\Theta}) \quad \xi(x_{t-1}, x_t) = p(x_{t-1}, x_t|z_{1t}^T, U_{1t}^T, \hat{\Theta}) \tag{8}
\]

where \( \hat{\Theta} \) is the latest estimate of \( \Theta \) representing all model parameters. The expectation of the observation likelihood can be written as

\[
Q(\Theta, \hat{\Theta}) = \sum_{t=1}^{T} \sum_{k=1}^{K} \sum_{j=1}^{K} x_{t-1,j} x_{t,k} ln A_{jk} + \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma(x_{t,k}) ln p(z_{t}|x_{t,k}),
\]
where \( p(z_t|x_{t,k}) \) is the observation model and \( K \) is the number of states that the latent variable can take and in our case is 5.

During the process of training the model parameters using the generalized EM algorithm, the sum of the expectation of the likelihood in equation (9) for all the \( P \) short-observation sequences of the current cell needs to be maximized with respect to \( \Theta \). In our method, the transition probabilities are parameterized through multiple layers of neural units, which has the effect of making the learning process smooth, as verified in [10].

**B. Modeling Long-term Dependency by Hierarchical Structure with Starting Place**

In this section we model how the local moving tendency from the current cell is dependent on initiating events happening many cells away at an earlier time by capturing the implicit long-term statistical correlations. The connection between higher level occurrences and cell-level movements is explored by introducing the concept of a starting place into the IOHMM forming a hierarchical input structure.

1) **Starting place**: From the macroscopic point view, human living environments are structural and functional. For example, in an office building, a corridor connects many rooms which can be offices, kitchens or bathrooms. These rooms, where people stop and stay, are typical examples of the ‘resting place’ [9]. The functional characteristics of the rooms decide about the macroscopic behavior of people in that the resting places are the origin and destination of human activities in the environment. The topological behaviors from one resting place to another globally characterize detailed cell-level movements. We believe that the resting place where the movement starts carries valuable information of long-term dependency of the local HMM process in a particular cell. We refer to it as the ‘starting place’ and incorporate it into our model.

As illustrated by the example in Figure 4, if a person starts from the resting place, office A, and enters the cell ‘C’ from ‘L’, then he/she is very likely to go into the kitchen, and thus exiting ‘C’ to ‘D’. On the other hand, if a person enters the cell ‘C’ again from ‘L’, but starting from a resting place somewhere close to the left-hand end of the corridor, then he/she is very likely to go through to the right-hand end of the corridor, and thus exiting ‘C’ to ‘R’. In the latter case, the index of the starting place is not restricted to rooms but extended to certain regions in the corridor.

![Fig. 4. An example of the one-step movement with the starting resting place in a corridor environment.](image)

2) **IOHMM with Hierarchical Input Structure**: The input \( U_{t+1} \) of our IOHMM based model is local in the sense that it varies for each short-observation sequence and that it varies in the same HMM process at different time steps. On the other hand, the starting place is used as a global input such that its value remains the same for all short-observation sequences that originate from that starting place. Figure 5 illustrates the hierarchical input structure involving the starting place, which is denoted as \( x^s \), and \( U_{t+1} \).

![Fig. 5. Our IOHMM based model with the hierarchical input structure.](image)

In a similar manner to construct the local input \( U_{t+1} \), the global input, \( x^s \), is coded by a number of binary variables corresponding to the set of selected starting places in the environment. With the starting place, the transition probabilities become

\[
p(x_{t+1}|x_t, U_{t+1}, x^s, W) \tag{10}
\]

3) **Training with Starting Place**: With the additional global input, \( x^s \), the mapping from input to transition probabilities in our IOHMM based model needs to be modified. We introduce a switching function \( g_j(x^s, w_j^{(1)}) \) into the mapping process, where \( j \) is the index of the starting place and \( w_j^{(1)} \) represents the weight for the corresponding starting place. By multiplying this function to the first layer variables of the neural network, the training data are grouped according to the starting place resulting in one model for each a starting place.

Suppose there are \( N \) selected starting places in the environment and they are coded by the members in the set \( \{x^{s_j} : j = 1, \ldots, N\} \). The first step of the neural network in equation (4) is changed to

\[
a_j = (\sum_{i=1}^{4} w_{j1}^{(1)} u_i^{(1)} + w_{j0}^{(1)}) g_j(x^s, w_j^{(1)}), \tag{11}
\]

and the switching function \( g_j(.) \) is defined as

\[
g_j(x^s, w_j^{(1)}) = \begin{cases} w_j^{(1)} & \text{if } x^s = x^{s_j} \\ 0 & \text{otherwise} \end{cases} \tag{12}
\]

where \( j = 1, \ldots, N \). The number of \( a_j \) in the neural network is decided by the number of selected starting places. The function \( g_j(.) \) acts as a switch in that if and only if \( x^s=x^{s_j} \) then \( a_j \) results in a non-zero value. Thus the training data are divided into \( N \) groups according to the starting place, and each \( w_j^{(1)} \) only applies to the corresponding group. However, it is worth noting that all other parameters of the neural network are shared by all training sequences.
The training procedure that is described in Section III-A.4 remains unchanged in all other aspects.

IV. INFERENCE OF MULTI-SCALE CONDITIONAL TRANSITION MAP

The overall transitional properties for the entire space that are captured in the MTCMap reflect the object moving trend in the environment. In this section, we show that knowledge of this model enables inferring future occurrence probabilities based on two initial observations.

We assume that two consecutive initial observations of the object in two adjacent cells are available and use this as the prior of the object location. The variables representing locations of the object at each time step form a directed graph with the next location variable conditioned on the previous one. We use the ancestral sampling approach [15] to sample the underlying joint distribution for the inference. The transition probabilities that are needed for the sampling are calculated from the learned model. In this process, as observations of the object location along the sampling path are not available, we use the prediction of occupancy of the neighbor cells as input to our IOHMM based model of the current cell for calculating the transition probabilities.

These inferred future occurrence probabilities provide valuable domain knowledge about where dynamic objects are potentially moving in the environment and can therefore be used to improve mobile robot path planning. For typical robotic navigation tasks from one location to another, the robot is provided with basic domain information such as obstacle locations, for example by a SLAM component [19]. The additional domain knowledge inferred from the MTCMap can be readily incorporated into path planning by using a popular informed incremental heuristic search algorithm for example D* Lite [20] to avoid collision with people and achieve the minimum-cost path or to maximize the chance of meeting people (for example in delivering coffee) and achieve the maximum-gain path.

V. EXPERIMENTS

The goal of these experiments is to verify the capability of the method we propose to capture the spatial-temporal dynamics of human movements. Considering the size of a human body, we choose a coarse representation of the environment by setting the grid size as 0.3 x 0.3m in our map. The experimental results show that this representation is sufficient for studying the motion patterns of people in our experimental environment - a corridor in an office building.

A. Experimental setup

We used a SICK LMS200 laser range-finder for collecting 2D data at approximately 37Hz. The environment is shown in Figure 6 and the arrows indicate certain interesting locations which are helpful to understand the results. The doors of an office, the elevator and the kitchen right next to the stairs are indicated by the blue (number 1), yellow (number 3) and green (number 4) arrows, and the doorway to the bathroom is indicated by the red arrow (number 2). We set up the laser range-finder at one end of the corridor and started data logging at 10:04. The data collection process went on till 20:23, corresponding to 10 hours and 19 minutes of laser scans.

B. Generating short-observation sequences

The endpoint model [21] is applied to generate occupancy observations for each cell in the map. Cells along the laser ray are treated equally and considered as ‘free’, and the cell where the laser ray ends is considered as ‘occupied’. These observations form sequences for all cells in the map.

The training of our IOHMM based model requires short-observation sequences due to the special left-to-right structure. Our model only focuses on the tendency of the object moving from one cell to another while not caring about fine motion details. With temporal and spatial discretization, we transform the original occupancy observation sequences into events of entering and occupying a grid cell. Then a set of short-observation sequences each in the form of \((x_1^T, U_1^T, x^s)\) are generated. Tracking of the starting place for each sequence is performed by analyzing the whole trajectory. We noticed in the experiments that the training process is significantly sped up due to the extraction of interesting events, especially considering that those events happen in a comparatively occasional manner during the prolonged data collection period.

C. Training Results

The macroscopic models are overlaid with the floor plan of the building and shown in 7. The model from training without the starting place is shown in Figure 7(a). The training with four starting places (two offices, the kitchen and the bathroom) results in four models, one for each starting place and Figure 7(b) shows the one for the kitchen. The arrows indicate moving tendency from the cell where the arrow starts to the cell where it points to. The length of arrows is proportional to the value of the corresponding transition probability that is calculated from the learned model parameters. In our current implementation, no training is carried out for cells on the edge of the map, as they have one or two neighbor cells missing, hence no arrows in them.

As shown by Figure 7(a), in most cells, arrows along the corridor are longer than sideways ones, which is in accordance with the way people mostly walk in a corridor. On the other hand, sideways arrows do appear in significant
This is because the modeled dynamics in each cell is comparatively spread out and certain extent of randomness exit at the bottom. The moving tendency in Figure 8(a) is move in the corridor. For example, a person walking out of the corridor on the top and bottom.

D. Inference of Future Occurrence Probabilities

In this experiment, we specify the two consecutive initial observations in two adjacent cells right outside the kitchen and repeat sampling the joint distribution 1000 times following the procedure described in Section IV. At each cell, the occupancy events are summed up for all samples. Figures 8(a) and 8(b) show the inference using the learned models shown in 7(a) and 7(b) respectively. In both cases, the tendency of movement is in accordance with how people move in the corridor. For example, a person walking out of the kitchen tends to go to the bathroom, some offices or the exit at the bottom. The moving tendency in Figure 8(a) is comparatively spread out and certain extent of randomness can be seen as some samples illustrate circling behaviors. This is because the modeled dynamics in each cell is a summary of human movements originating from all resting places. In contrast, the moving tendency in Figure 8(b) is more focused and the prediction of the future trajectory is more informative due to the guidance of the starting place by way of long-term dependencies. The tendency in Figure 8(b) shows that more people move upwards after exiting the kitchen, which is consistent with the fact that there are more offices towards the end of the corridor on the top (beyond the figure).

E. Evaluation of Inference with Ground Truth

We further evaluate the accuracy of the inference from the model shown in Figure 7(b) with the starting place being the kitchen. True human tracks are extracted from the collected data and those starting from the kitchen are used as the ground truth. The resting places that these tracks are heading are designated beforehand. In each trial, one ground truth track is picked up and a trajectory is generated with the learned model and initial observations. The success is decided by comparing the predicted destination resting place with that of the ground truth track. We repeat this process for the whole set of ground truth tracks according to different length of the observed partial track and for each length we perform the prediction 1000 times. Figure 9 shows the success rate w.r.t. the length of the observed partial track. This length is in terms of the number of cells traversed. We can see that as the length of observed track increases the
ability to predict the correct destination increases and with an observed track length of approximately 2m we predict all destination correctly for the test data.

Fig. 9. Success rate of trajectory prediction against the length of the partial track that is observed in terms of the number of cells traversed.

VI. CONCLUSIONS

In this paper we presented the Multi-scale Conditional Transition Map, a novel method for modeling the dynamics of moving objects, such as humans, in the environment with a grid-based representation. In contract to existing methods, the transition model in our approach reflects both local correlations and long-term dependencies. For each grid cell, we employ a variant of the left-to-right HMM to explicitly model the local moving tendency from the current cell to a neighboring cell. By exploiting the IOHMM formulation, we make the transition probabilities adaptive to the entering direction, capturing the spatial-temporal correlations across the three cells involved. On the higher level of the input hierarchy of the IOHMM, we introduce the starting place, capturing the long-term dependency of local activities on faraway initiating events. Close examination shows that the IOHMMs in different cells are independent conditioned on the observation set, which facilities the training process.

The method we propose seamlessly incorporates information from both the cell and topology levels into local IOHMM processes for all grids in the environment. The transition models collectively enable inference of future occurrence probabilities of moving objects through the spatial and temporal space, which can be readily used for path planning. Our method is based on regularities in people’s movements in human-living environments. The capability of capturing the ‘regular tracks’ and utilizing the information contained in the ‘habitual’ behavior is valuable in various robotic tasks such as long-term mapping and/or localization with dynamic objects and navigation involving avoiding or interacting with people.

ACKNOWLEDGMENT

This work was funded by SSF through its Centre for Autonomous Systems and the EU FP7 project STRANDS (600623).

REFERENCES