Modeling Spatial-Temporal Dynamics of Human Movements for Predicting Future Trajectories

Zhan Wang and **Patric Jensfelt** and **John Folkesson**
Computer Vision and Active Perception Lab
KTH Royal Institute of Technology
{zhanw,patric,johnf}@kth.se

Abstract

This paper presents a novel approach to modeling the dynamics of human movements with a grid-based representation. 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 a higher level, we introduce the place where the whole trajectory originated into the IOHMM framework forming a hierarchical input structure. Therefore, we manage to capture both local spatial-temporal correlations and the long-term dependency on faraway initiating events, thus enabling the developed model to incorporate more information and to generate more informative predictions of future trajectories. The experimental results in an office corridor environment verify the capabilities of our method.

Introduction

For robots engaged in long-term operations in real world scenarios, one of the major challenges is to learn a model of the environment with dynamic objects, such as people. In modeling dynamics, many methods employ the occupancy grid map representation (Moravec and Elfes 1985) for mobile robots (Saarinen, Andreasson, and Lilienthal 2012; Saarinen et al. 2013; Arbuckle, Howard, and Matari 2004; Meyer-Delius, Beinhofer, and Burgard 2012). In the work (Meyer-Delius, Beinhofer, and Burgard 2012), 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. However, the work assumes that the dynamics in the environment is due to a stationary process, relying on the online training procedure to handle non-stationary dynamics. The recent work (Wang et al. 2014), which is based on the Input-Output HMM (IOHMM), is capable of modeling the inhomogeneous motion patterns of dynamic objects and responding to high dynamics without relying on online learning (Cappe 2011) or recency-weighted (Biber and Duckett 2009) techniques.

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. Locally, 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 (Wang et al. 2014), but by way of the occupancy of cells rather than explicitly modeling the tendency of moving. On a larger scale, 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’ (Bennewitz et al. 2005), for example offices, kitchens and bathrooms. The events in high-level human behaviors, such as leaving a resting place, obviously have significant influence on the local movements in a specific location. 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.

In this paper we propose a method to model the dynamics of human movements by capturing both local spatial-temporal correlations and implicit long-term dependencies with a grid-based representation of the environment. 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 (Bengio and Frasconi 1996; Bengio 1999). 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. The two main contributions of our work are:

Firstly, we present a method that is capable of capturing local spatial-temporal correlations in dynamics of human movements. The developed model, which is based on the inhomogeneous IOHMM, provides rich dynamics to accommodate the ever-changing nature of the motion pattern of dynamic objects, and thus being capable of capturing local correlations more efficiently.

Secondly, we provide a way of incorporating long-term dependency on faraway initiating events into the above IOHMMs and thus enable these processes collectively to generate more instructive and informative prediction of future moving trajectories.

Our method works by learning cell transitions dependent on both the adjacent grid cell that was occupied just before the currently occupied cell was filled (i.e. the local direction) and the high-level starting place of the trajectory (i.e. the global source).

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. (Arbuckle, Howard, and Matari 2004) 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. Biber et al. (Biber and Duckett 2009) developed a multiple timescale map representation using sample-based techniques. The map of each layer is constructed from a sample set of sensor data used as primitives. Each set is updated at a varied learning rate to adapt to different extent of dynamics in the environment according to the timescale. Saarinen et al. (Saarinen et al. 2013) 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 (Meyer-Delius, Beinhofer, and Burgard 2012), 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. However, the method is inherently homogeneous averaging dynamics in a certain timescale. The online training procedure provides certain adaptive capabilities from the recent tendency. In (Saarinen, Andreasson, and Lilienthal 2012), 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.

Common to the above methods based on occupancy grids is that each individual cell is modeled independently. In the Conditional Transition Maps developed by Kucner et al. (Kucner et al. 2013), 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. In the recent work (Wang et al. 2014), the IOHMM is employed to model the occupancy of each cell. The observations of neighboring cells are used as input to the IOHMM of the current cell, and hence local spatial correlations are captured. The transition parameters are estimated by a training procedure using generalized EM within the IOHMM framework. However both these two methods only capture local correlations, while in the work we propose in this paper both local spatial-temporal correlations and long-term dependencies are captured, resulting in a more informative model.

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.

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.

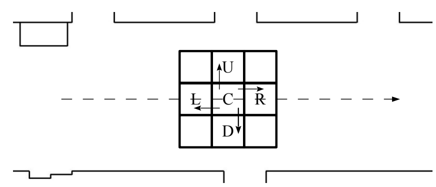

Figure 1: An example of one-step movement in a corridor environment, moving out of the current cell C to R, with the entering cell being L, as indicated by the dashed arrowed line. The solid arrowed lines indicate possible moving directions from the current cell C.

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 cells involved. Details of these two aspects of our model will be described in the following two subsections.

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.

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(a) shows the transition diagram. The corresponding 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.

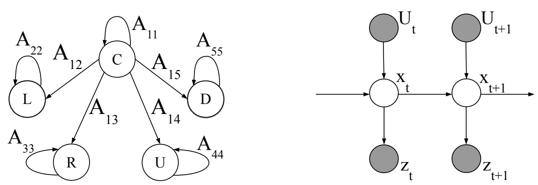

Figure 2: Two aspects of our model of the one-step movement. (a) Transition diagram of the variant of the left-to-right HMM. (b) Structure of the IOHMM.

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 in 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 2(b) 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 current IOHMM 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.

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 (1) 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 the next section.

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.

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 $A_{t+1,lk} = p(x_{t+1,k}|x_{t,l}, U_{t+1})$, in which $x_{t+1,k}$ 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, $z^{n1}_{t-1}, z^{n2}_{t-1}, z^{n3}_{t-1}, z^{n4}_{t-1}$. The input at time step $t+1$ is $U_{t+1} = [z^{n1}_{t-1}, z^{n2}_{t-1}, z^{n3}_{t-1}, z^{n4}_{t-1}]$, where z^{ni}_{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 as follows (Bishop 2006). For clarity purpose, the time subscript is omitted in the rest of this subsection and the input U_{t+1} is 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

$$a_j = \sum_{i=1}^{4} \rho^{(1)}_{ji} u^i + \rho^{(1)}_{j0}, \tag{1}$$

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

Then each of a_j, known as activation, is transformed by using a sigmoidal function $b_j = h(a_j)$ to obtain the hidden units, b_j, where $j = 1, \ldots, 4$.

Lastly the hidden units are linearly combined to obtain the transition probabilities $A_{1k} = \sum_{j=1}^{4} \rho^{(2)}_{kj} b_j + \rho^{(2)}_{k0}$, where $k = 1, \ldots, 5$. The weights and biases are collectively denoted as $\rho = [\rho^{(1)}_{ji}, \rho^{(1)}_{j0}, \rho^{(2)}_{kj}, \rho^{(2)}_{k0}]$, where $i, j = 1, \ldots, 4$, $k = 1, \ldots, 5$.

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 (Bengio and Frasconi 1996) with the generalized EM algorithm (Dempster, Laird, and Rubin 1977) 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 (Levinson, Rabiner, and Sondhi 1983).

- As shown by the transition diagram in Figure 2(a), the process always starts from the first state, C, termed as the starting state, and hence the prior is set as $\pi = (1, 0, 0, 0, 0)$ and not reestimated.
- At the beginning of the forward-backward scheme (Rabiner 1989), set the transition probabilities $A_{lk} = 0, (l > k)$ and $A_{lk} = 0, (l < k, l > 1)$.
- The probability $\beta(x_t)$ in the forward-backward scheme is defined as $\beta(x_t) = p(z_{t+1}, \ldots, z_T | x_t, U_{t+1}, \ldots, U_T)$, $(t = 1, \ldots, T-1)$, where T is the total number of time steps in the current training sequence (Bengio and Frasconi 1996). The initial condition $\beta(x_T)$ is set as

$$\beta(x_T) = \begin{cases} 1 & \text{if } x_T \in \{L, R, U, D\} \\ 0 & \text{if } x_T \in \{C\}. \end{cases} \tag{2}$$

Suppose the training data for the current cell are a set of P short-observation sequences, $\{(z_1^{T_p}(p), U_1^{T_p}(p)); p = 1, \ldots, P\}$. Each sequence consists of all observations $z_1^{T_p}(p) = \{z_1(p), \ldots, z_{T_p}(p)\}$ and inputs $U_1^{T_p}(p) = \{U_1(p), \ldots, U_{T_p}(p)\}$ 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, $\gamma(x_t) = p(x_t | z_1^T, U_1^T, \hat{\Theta})$, and $\xi(x_{t-1}, x_t)$ denote the joint posterior distribution of two successive latent variables, $\xi(x_{t-1}, x_t) = p(x_{t-1}, x_t | z_1^T, U_1^T, \hat{\Theta})$, where $\hat{\Theta}$ is the latest estimate of Θ representing all model parameters. The expectation of the observation likelihood can be written as

$$\begin{aligned} Q(\Theta, \hat{\Theta}) \quad = \quad & \textstyle\sum_{t=2}^{T} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi(x_{t-1,j}, x_{t,k}) ln A_{jk} \\ & + \textstyle\sum_{t=1}^{T} \sum_{k=1}^{K} \gamma(x_{t,k}) ln p(z_t | x_{t,k}), \end{aligned} \tag{3}$$

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 (3) for all the P short-observation sequences of the current cell needs to be maximized with respect to Θ. 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 (Bengio and Frasconi 1996).

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.

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. 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. Therefore, these topological behaviors from one resting place to another globally characterize detailed cell-level movements. It is the resting place where the movement starts that carries valuable information of long-term dependency of the local HMM process in a particular cell under concern, which we refer to as the starting place.

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 3 illustrates the hierarchical input structure involving the starting place, which is denoted as x^s, and U_{t+1}.

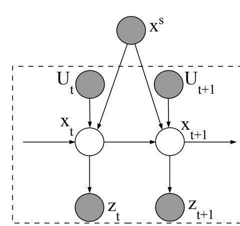

Figure 3: Our IOHMM based model with the hierarchical input structure.

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 $p(x_{t+1} | x_t, U_{t+1}, x^s, \rho)$.

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 $g_j(x^s, \rho_j^{(1)})$ into the mapping process, where j is the index of the starting place and $\rho_j^{(1)}$ 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 a number of models each for a starting place.

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

$$a_j = (\sum_{i=1}^{4} \rho_{ji}^{(1)} u^i + \rho_{j0}^{(1)}) g_j(x^s, \rho_j^{(1)}), \tag{4}$$

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

$$g_j(x^s, \rho_j^{(1)}) = \begin{cases} \rho_j^{(1)} & \text{if } x^s = x^{sj} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

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^{sj} 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 $\rho_j^{(1)}$ 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 the previous section remains unchanged in all other aspects.

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 $0.3 \times 0.3m$ 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.

We used a SICK LMS200 laser range-finder for collecting 2D data at approximately 37Hz. 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.

The endpoint model (Thrun, Burgard, and Fox 2006) is applied to generate occupancy observations for each cell in the map. 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 (z_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.

Training Results

The macroscopic models from our training results produced in an offline manner using the collected data, which are overlaid with the floor plan of the building, are shown in Figure 4. The model from training without the starting place is shown in Figure 4(a). The training with four starting places (two offices, the kitchen and the bathroom) results in four models each for a starting place and Figure 4(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. 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.

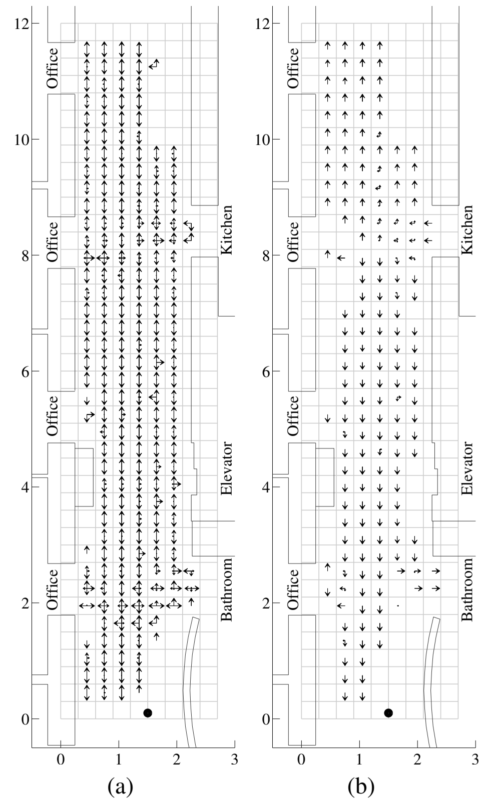

Figure 4: Resulting models from (a) training without starting places, and (b) training with four starting places (the one with the kitchen being the starting place is shown), which are overlaid with the floor plan of the building. The length of arrows is proportional to the value of the corresponding transition probability that is calculated from the learned model parameters in ρ and very weak transitions are not displayed. The big black dot at the bottom indicates the sensor location.

As shown by Figure 4(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 magnitude in areas right outside the office, the elevator, the kitchen and the doorway to the bathroom where people walk sideways entering or leaving those places. Figure 4(b) clearly illustrates the tendency of people who exit from the kitchen heading to the bathroom, nearby offices, or the two ends of the corridor on the top and bottom.

Prediction of Future Trajectories

The overall transition model for the entire map, which is represented by the collection of parameters in ρ for all cells, reflects the object moving trend in the environment. In this section, we show that knowledge of this model enables inferring future trajectories 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 (Bishop 2006) to sample the underlining joint distribution and generate the trajectory. The transition probabilities that are needed for the sampling are calculated from the learned model as illustrated in Figure 4. In this process, as observations of the object location along the trajectory 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.

In the experiment, we specify the two consecutive initial observations in two adjacent cells right outside the kitchen and repeat sampling the joint distribution 1000 times. At each cell, the occupancy events are summed up for all samples. Figures 5(a) and 5(b) show the predicted trajectories using the learned models shown in 4(a) and 4(b) respectively. In both predictions, 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 5(a) is comparatively spread out and certain extent of randomness can be seen as some samples illustrate behaviors such as circling. 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 5(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.

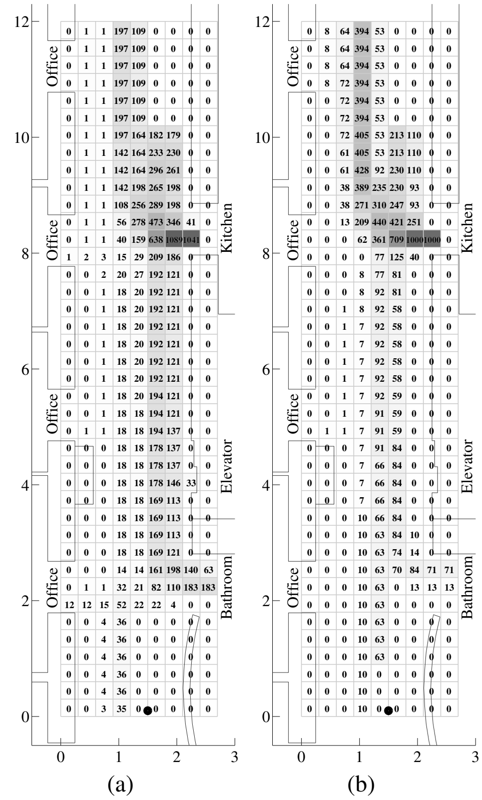

Figure 5: Results of prediction with the initial observations being right outside the kitchen using the learned models shown in Figure 4 from (a) training without the starting place, and (b) training with four starting places (the model for the kitchen is used). The number of occupancy events for each cell is written in the cell and indicated by the darkness of the cell (the darker the more occupancy events).

Conclusions

In this paper we proposed a novel method for modeling the dynamics of human movements in the environment with a grid-based representation. 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.

Our approach has two advantages over the state-of-the-art. Firstly, being an inhomogeneous model based on the IOHMM, it provides rich expressive power to accommodate the inherently inhomogeneous nature of the dynamics of moving objects, and thus capture local spatial-temporal correlations more efficiently. Secondly, for each local process the long-term dependency on faraway initiating events is modeled by introducing the starting place and thus collectively enabling more instructive and informative prediction of future trajectories.

Our method is based on regularities in people's movements in human-living environments. For example, in an office building, people normally walk straight along the corridor and only turn sideways in cases of entering rooms such as a kitchen or an office. On the higher level, people exiting a certain room tend to visit some rooms more often than others, subjective to the structural and functional design of the environment. The method we propose seamlessly incorporates information from the cell and topology levels into local IOHMM processes for all cells in the map, which then collectively are able to produce predictions of local immediate movements and global long-distance trajectories. This capability of capturing the 'regular tracks' and utilizing the information contained in the 'habitual' behavior is obviously valuable in various robotic tasks such as long-term mapping and/or localization with dynamic objects and navigation involving avoiding or interacting with people.

Acknowledgments

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

References

Arbuckle, D.; Howard, A.; and Matari, M. 2004. Temporal occupancy grids: a method for classifying the spatio-temporal properties of the environment. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 409–414.

Bengio, Y., and Frasconi, P. 1996. Input-Output HMMs for sequence processing. *IEEE Trans. on Neural Networks* 7(5):1231–1249.

Bengio, Y. 1999. Markovian models for sequential data. *Neural Comput. Surv.* 2:129–162.

Bennewitz, M.; Burgard, W.; Cielniak, G.; and Thrun, S. 2005. Learning motion patterns of people for compliant robot motion. *The International Journal of Robotics Research* 24(1):31 – 48.

Biber, P., and Duckett, T. 2009. Experimental analysis of sample-based maps for long-term SLAM. *The International Journal of Robotics Research* 28(1):20–33.

Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*. Springer.

Cappe, O. 2011. Online EM algorithm for Hidden Markov Models. *Journal of Computational and Graphical Statistics* 20(3):728–749.

Dempster, A. P.; Laird, N. M.; and Rubin, D. B. 1977. Maximum-likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39(1):1–38.

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

Levinson, S. E.; Rabiner, L. R.; and Sondhi, M. M. 1983. An introduction to the application of the theory of probabilistic functions of a markov process to automatic speech recognition. *Bell System Technical Journal* 62(4):1035 – 1074.

Meyer-Delius, D.; Beinhofer, M.; and Burgard, W. 2012. Occupancy grid models for robot mapping in changing environments. In *AAAI conference on artificial intelligence*.

Moravec, H., and Elfes, A. 1985. High resolution maps from wide angle sonar. In *IEEE Int. Conf. Robotics and Automation (ICRA)*, 116–121.

Rabiner, L. R. 1989. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2):257–285.

Saarinen, J.; Andreasson, H.; and Lilienthal, A. J. 2012. Independent markov chain occupancy grid maps for representation of dynamic environment. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3489–3495.

Saarinen, J. P.; Andreasson, H.; Stoyanov, T.; and Lilienthal, A. J. 2013. 3D normal distributions transform occupancy maps: an efficient representation for mapping in dynamic environments. *The International Journal of Robotics Research* 32(14):1627–1644.

Thrun, S.; Burgard, W.; and Fox, D. 2006. *Probabilistic Robotics*. MIT Press.

Wang, Z.; Ambrus, R.; Jensfelt, P.; and Folkesson, J. 2014. Modeling motion patterns of dynamic objects by IOHMM. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.