Modeling motion patterns of dynamic objects by IOHMM

Zhan Wang[1], Rares Ambrus[1], Patric Jensfelt[1] and John Folkesson[1]

Abstract— This paper presents a novel approach to model motion patterns of dynamic objects, such as people and vehicles, in the environment with the occupancy grid map representation. Corresponding to the ever-changing nature of the motion pattern of dynamic objects, we model each occupancy grid cell by an IOHMM, which is an inhomogeneous variant of the HMM. This distinguishes our work from existing methods which use the conventional HMM, assuming motion evolving according to a stationary process. By introducing observations of neighbor cells in the previous time step as input of IOHMM, the transition probabilities in our model are dependent on the occurrence of events in the cell's neighborhood. This enables our method to model the spatial correlation of dynamics across cells. A sequence processing example is used to illustrate the advantage of our model over conventional HMM based methods. Results from the experiments in an office corridor environment demonstrate that our method is capable of capturing dynamics of such human living environments.

I. INTRODUCTION

For robots operating in real world environments, one of the major challenges is to deal with dynamic objects, such as people and vehicles. In such scenarios, the capability of learning the dynamic patterns of the objects is highly relevant to the success of the tasks. Different methods have been developed to tackle this problem based on various techniques, such as [1], [2].

Occupancy grids [3] are one of the most popular representations of environments where mobile robots operate, and are also used for dynamic environment studies [4] [5] [6] [7]. In a very recent work, Meyer-Delius et al. [8] generalized the occupancy grid map and released the static environment assumption by introducing the hidden Markov model (HMM). The generalized occupancy grid map explicitly accounts for changes in the environment and exhibits excellent capability to distinguish between highly dynamic (such as people), semi-dynamic (such as chairs) and static objects.

Although the HMM has been shown to be a powerful tool in dynamic occupancy grid mapping [8] [9], the transition parameters in conventional HMMs are fixed after training, which results in a homogeneous Markov chain. However, dynamic objects, such as pedestrians or vehicles, usually move in a manner that the motion pattern changes from time to time. The inherently inhomogeneous nature of moving behavior of these objects renders the conventional HMM a suboptimal model for scenarios with high dynamics. Online learning [10] or recency-weighted [1] techniques can be used to make parameters of HMM adaptive to the recent

tendency. However, inherently the model still remain homogeneous averaging dynamics in a certain timescale and will encounter difficulties in coping with the randomly changing nature of dynamics in the environment. For example, during lunch hours, the frequency of people entering the kitchen is often notably higher than other times if one looks over 30 minute timescales, whereas each individual entering event will happen randomly, which makes learning a pattern in the frequency at shorter timescales such as 5 minutes problematic. In this case, the conventional HMM will find it hard to predict what will happen next, even by averaging from the recent past. Therefore, the conclusion is that for tasks of capturing regular motion patterns of dynamic objects in the real world, the model will need to be inhomogeneous.

One observation we made is that the objects move in a continuous manner in the spatial dimension. That is, they do not appear in a particular location from nowhere or suddenly disappear, except at the edges of our environment such as an entry door. As a consequence, before they reach a certain spot, they have to enter some neighboring spot then move to the destination according to their motion pattern. Therefore, the occurrence of entering events in a neighboring region contains significant amount of information about possible future events in a region of interest. This observation motivates the proposed work which builds connections between past and future events across cells, and models certain spatial-temporal aspects of dynamics of objects.

In this paper, we propose a method to model dynamics in the environment by assigning an input-output hidden Markov model (IOHMM) to each occupancy grid cell. The IOHMM was originally introduced by Bengio et al. [11] [12] for sequence processing. An IOHMM is a HMM for which the observation and transition probabilities are conditional on the input sequence. We introduce observed events happening in neighboring IOHMMs as input to the current one, thus facilitate information flowing across IOHMMs which run in parallel for all cells, and enable incorporating precursor information from neighboring processes. This leads to two contributions of our work:

First, we improve individual IOHMM processes by the addition of external information from neighboring IOHMMs.

Second and more importantly, the model we develop is capable of capturing spatial correlation across different processes, and thus modeling the corresponding motion pattern by the parameters characterizing dependency of transition probabilities on the inputs.

[1]Zhan Wang, Rares Ambrus, Patric Jensfelt and John Folkesson are with Computer Vision and Active Perception Lab, Centre for Autonomous Systems, CSC, KTH Royal Institute of Technology, Stockholm 100 44, Sweden {zhanw,raambrus,patric,johnf}@kth.se

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. [7] 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. [1] 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. [6] 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.

Meyer-Delius et al. [8] developed the dynamic occupancy grid, which utilizes a HMM with a two-state Markov chain to model the state of a cell. The dynamics of each cell is explicitly represented by the transition probabilities. Both online and offline methods are provided for training the HMM. Introducing the HMM for modeling the grid cells successfully released the assumption of static environments for the traditional occupancy grid map representation. However, this work assumes that the dynamics in the environment is due to a stationary process, although the online training procedure provides certain adaptive capabilities from the recent tendency to handle non-stationary dynamics. 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.

So far all above methods that are based on occupancy grids have modeled each individual cell independently. In the Conditional Transition Maps developed by Kucner et al. [4], the cross-cell spatial relation is modeled as a probability distribution of an object leaving to a neighbor cell conditioned on which cell it comes from. The cross-correlation is used to learn the transition parameters. This work is the closest to our method. They assume that objects move one cell each time step by proper discretization in space and time domains, and the value of conditional transition parameters are learned by counting events. In our method, the parameters are estimated by a training procedure using generalized EM within the IOHMM framework, and we do not require that objects move at a constant speed even in a discrete manner.

III. MODELING FOR DYNAMIC ENVIRONMENTS BY IOHMM

A. Occupancy grid map representation

We divide the overall map into occupancy grid cells, which are denoted as $C = \{c^1, c^2, \ldots, c^M\}$, where M is the total number of cells. For each cell, let x_t be the binary random variable representing the occupancy state at time t, $x_t \in \{occupied, free\}$. The observation for each cell at time t, z_t, is also a binary random variable with two states, $\{occupied, free\}$. Throughout this paper, indexes of cells are denoted by superscripts and time steps are denoted by subscripts for occupancy states and observations.

In our work, each cell is modeled using an IOHMM, which is introduced in the next section. More details about the IOHMM can be found in [11].

B. IOHMM

Hidden Markov models or HMMs [13] are statistical models of sequential data. The latent variables, denoted by x_t, are discrete and form a discrete-state Markov chain as illustrated by the graph in Fig. 1. The observed variables, z_t, may be discrete or continuous.
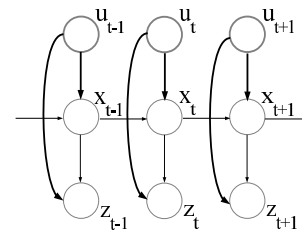

Fig. 1. An example of an IOHMM and an example of a HMM (part of the graph without darkened lines).

The IOHMM is an extension of the HMM, which has been successfully applied in areas such as audio-visual synthesis [14]. In an IOHMM we have a sequence of input variables, u_t, as in Fig. 1, whose values influence either the distribution of latent variables or observed variables, or both. This is illustrated by the arc from the input to the latent variable and the arc from the input to the observed variable in Fig. 1. By introducing the input sequence, the IOHMM provides the possibility of incorporating additional information into the process modeled by the conventional HMM.

C. Modeling dynamic environments using IOHMM

1) Main idea: We make the observation that objects move in a continuous manner. That is, before they move into the cell of concern, they have to enter the neighbor cell. Therefore, the events happening in neighbor cells provide valuable information about possible future event in the cell of concern. This observation motivates this work which tries to build connections between past events and future possible events across cells, and these connections across cells provide a convenient way of modeling certain aspects of dynamics of objects (spatial correlation of objects). To model these connections we do not need to include any dependency of the observation variable on the input as shown in the general IOHMM model of Fig. 1.

Based on above idea, we propose using the IOHMM structure as in the graph of Fig. 2 for our modeling. We let the occupancy state, represented by x_t, to be dependent on the input sequence, which is chosen to be the observations

of the neighbor cells at the previous time step. For example, the input U_t^m for the cell c^m consists of a set of observations, z_{t-1}^n, of the neighbor cells of c^m. In Fig. 2, U_t is used to denote the input which is a collection of quantities.
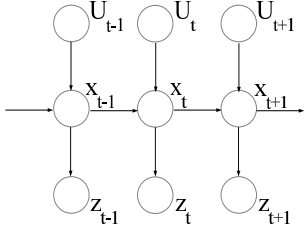

Fig. 2. The IOHMM for a single cell used in our method, with the superscript omitted. The input sequence, U_t, will in turn depend on past events in neighbor cells.

The occupancy state, x_t, depends on the input by way of the transition model, which specifies the probabilities of a cell to transfer from one state to another. We believe that the transition probabilities are highly related to the manner by which the cells are occupied in sequence, and thus can be used to capture the motion pattern of objects in the environment. As such, our IOHMM based model makes the transition probabilities dependent on the observations of neighbor cells. The details are described in the following subsection.

2) Mapping input to transition using neural network: The state transition matrix consists of a set of transition probabilities. Each element is a conditional distribution, through which the probability distribution of x_t depends on the state of the occupancy state variable at the previous time step, x_{t-1}, and has the form of

$$A_{t,jk} = p(x_{t,k}|x_{t-1,j}), \tag{1}$$

in which $x_{t,k}$ denotes the occupancy state variable at time t being in the kth state. We assume that the transition probabilities are changing with time. In our case, the occupancy state variable may only take two states, $\{occupied, free\}$, so the transition probabilities constitute a two by two transition matrix,

$$\begin{bmatrix} 1 - A_{12} & A_{12} \\ A_{21} & 1 - A_{21} \end{bmatrix}, \tag{2}$$

where the time subscripts are omitted

We currently use observations from four neighbors (left, right, up and down) as input due to the relatively large grid size (0.3m as used in the experiments in Section IV). The input for the time step t is

$$U_t = [z_{t-1}^{n1}, z_{t-1}^{n2}, z_{t-1}^{n3}, z_{t-1}^{n4}], \tag{3}$$

where z_{t-1}^{ni} represents the observation of the ith neighbor cell at time $t-1$. Please note that when the grid size is small, the chances of objects moving diagonally increase and observations from all eight neighbor cells should be used as input in equation (3).

Depending on the input, with U_t being denoted as $U_t = [u_t^1, u_t^2, u_t^3, u_t^4]$, the transition probabilities are formulated using a two layer neural network, defined as follows [15].

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

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

where $j = 1, \ldots, 4$, and the superscript (1) indicates the first layer of the network. The parameters $\rho_{ji}^{(1)}$ and $\rho_{j0}^{(1)}$ 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

$$A_{12} = \sum_{j=1}^{4} \rho_{1j}^{(2)} b_j + \rho_{10}^{(2)}, \tag{6}$$

$$A_{21} = \sum_{j=1}^{4} \rho_{2j}^{(2)} b_j + \rho_{20}^{(2)}, \tag{7}$$

where A_{12} represents the state transition probability from 'occupied' to 'free', and A_{21} represents the opposite. The weights and biases are collectively denoted as $\rho = [\rho_{ji}^{(1)}, \rho_{j0}^{(1)}, \rho_{1j}^{(2)}, \rho_{10}^{(2)}, \rho_{2j}^{(2)}, \rho_{20}^{(2)}]$, where $i, j \in \{1, 2, 3, 4\}$.

Taking into account the dependence on the input, the transition probabilities have the form
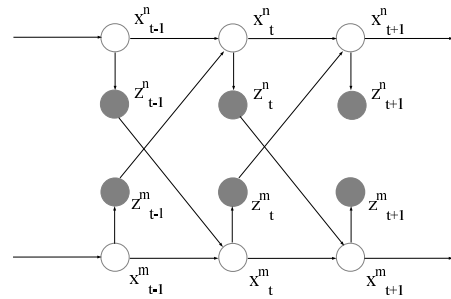
$$p(x_t|x_{t-1}, U_t, \rho). \tag{8}$$

Fig. 3. An example of our model using the IOHMM with two chains.

3) Analysis of conditional independence properties: The analysis is carried out by using the example with two Markov chains as shown in the graph of Fig. 3. The observed nodes are indicated with shaded circles. Clearly, all the paths connecting any latent variable in the m chain, for example, x_t^m, to any latent variable in the n chain, for example, x_{t+1}^n, are head-to-tail at the observation nodes, and the corresponding nodes are d-separated [15]. Therefore the two chains are independent conditioned on the observation set.

It is also easy to examine that the conditional independence property holds for the individual Markov chain by noting, for example, x_{t-1}^m and x_{t+1}^m are d-separated by x_t^m together with the observation nodes. These conclusions can be readily extended to a graph with any number of Markov chains.

Therefore, when EM is used for training parameters in our model, the exact E step can be carried out using the efficient forward and backward recursions along each individual Markov chain independently, leading to an efficient training procedure which will be described in Section III-C.5.

4) Observation probabilities and prior: The observation model in our case is discrete and the observation can only take two values, $\{occupied, free\}$. As shown in Fig. 2, in our model the observation probabilities are not dependent on the sequence of input, and are fixed. With the subscript t omitted, the observation probabilities are defined by

$$B_{jk} = p(z_k|x_j), \qquad (9)$$

in which z_k indicates the observation of the kth state and x_j indicates the occupancy state variable taking the jth state. The observation matrix consists of a set of such conditional probabilities

$$B = \begin{bmatrix} 1 - B_{12} & B_{12} \\ B_{21} & 1 - B_{21} \end{bmatrix}. \qquad (10)$$

The marginal distribution $p(x_1)$ for the initial occupancy state variable is represented by a vector π. This completes the definition of our model by IOHMM for the dynamics of each cell in the occupancy grid map, which requires the parameters $\Theta = \{\rho, B, \pi\}$.

5) Training procedure: In this section, we describe our method of training the parameters for each IOHMM. We generally follow the training scheme in [11] with the generalized EM algorithm [16]. The process of training the transition parameters, ρ, is novel in our method. The parameters to be learned are $\Theta = \{\rho, B, \pi\}$, and the procedures are as follows.

Let $\gamma(x_t)$ denote the marginal posterior distribution of the occupancy state variable x_t, and $\xi(x_{t-1}, x_t)$ denote the joint posterior distribution of two successive occupancy state variables, given the observations, the inputs, and the model parameters,

$$\begin{aligned} \gamma(x_t) &= p(x_t|Z, U_1, \ldots, U_T, \hat{\Theta}) \\ \xi(x_{t-1}, x_t) &= p(x_{t-1}, x_t|Z, U_1, \ldots, U_T, \hat{\Theta}) \end{aligned} \qquad (11)$$

where Z denotes the set of all observations, $\hat{\Theta}$ is the latest estimate of Θ and T is the total number of time steps. Then EM is to maximize the expectation of the observation likelihood

$$\begin{aligned} Q(\Theta, \hat{\Theta}) &= \sum_{k=1}^K \gamma(x_{1,k}) ln(\pi_k) \\ &+ \sum_{t=2}^T \sum_{j=1}^K \sum_{k=1}^K \xi(x_{t-1,j}, x_{t,k}) ln A_{jk} \\ &+ \sum_{t=1}^T \sum_{k=1}^K \gamma(x_{t,k}) ln p(z_t|x_{t,k}), \end{aligned} \qquad (12)$$

where K is the number of states that the occupancy state variable can take and in our case is 2.

- E step
  Using the estimate of ρ from the last iteration, the transition probabilities, A_{ij}, can be calculated based on the input values. For each state, the probabilities α and β, conditioned on the inputs, defined as

  $$\begin{aligned} \alpha(x_t) &= p(z_1, \ldots, z_t, x_t|U_1, \ldots, U_t) \\ \beta(x_t) &= p(z_{t+1}, \ldots, z_T|x_t, U_{t+1}, \ldots, U_T), \end{aligned} \qquad (13)$$

  can be calculated using A_{ij} and $\hat{\Theta}$, following the forward-backward scheme [17]. Thus the probabilities $\gamma(x_t)$ and $\xi(x_{t-1}, x_t)$ conditioned on the inputs can be calculated.

- M step
  In the M step, the parameters $\Theta = \{\rho, B, \pi\}$ are updated by maximizing the likelihood function in equation (12).
  The observation probabilities and prior can be updated as

  $$B_{ik} = \frac{\sum_{t=1}^T \gamma(x_{t,k}) z_{t,i}}{\sum_{t=1}^T \gamma(x_{t,k})}, \qquad (14)$$

  where $x_{t,k}$ indicates the occupancy state variable taking the kth state at time step t, $z_{t,i}$ indicates the ith state is observed at time step t, and

  $$\pi_k = \frac{\gamma(x_{1,k})}{\sum_{j=1}^K \gamma(x_{1,j})}. \qquad (15)$$

  Updating the parameters in ρ requires maximizing the second term on the right hand side of equation (12) with respect to ρ. 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 [11].

D. Illustration with an example

The model we propose is inherently an inhomogeneous HMM. In order to highlight the difference of our method and those based on the conventional homogeneous HMM, we provide a simple example of sequential data processing.

The observation sequences are shown in Fig. 4(a), corresponding to what will be observed for one cell and its four neighbors in a occupancy grid map, where $'0'$ and $'1'$ correspond to 'free' and 'occupied' respectively. The data of the sequence in concern (current cell) is used for the HMM training by EM, and the data of all sequences are used to train our IOHMM based model using the procedure described in Section III-C.5. Only the transition probability parameters are trained, which in our case are represented by ρ. Then the training results are applied to the same sequences for estimation. The observation model used for both methods is $p(z_i|x_i) = 0.9, i = 1, 2$.

It is important to note that in our method, the training and estimation start from time step 2 in Fig. 4(a). This is because the observations of the previous time step in the other four sequences are used as input to the sequence in
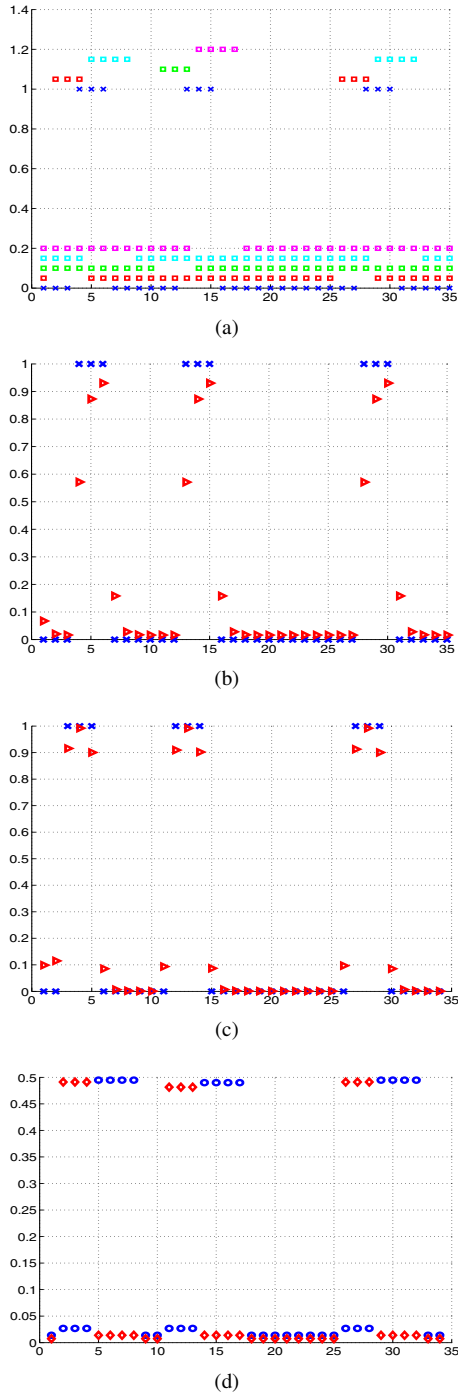
Fig. 4. Observations, training results and estimates of the example. The binary observations (either 1 or 0) are shown in (a). Data of the sequence in concern are indicated by blue crosses, while data of other related sequences by squares, each lifted a small interval to avoid overlapping. Estimates of occupancy probability from the HMM are shown in (b) with red triangles while data are displayed with blue crosses. Estimates of our method are shown in (c) and the transition probabilities are shown in (d), A_{21} with red diamonds and A_{12} with blue circles.

concern at the current time step. The estimation results in Fig. 4(b) and 4(c) show that our model is overall superior to that based on the HMM, and especially it is more reactive for changing of states. For example, at time step 4 in Fig. 4(b) when the observation is 'occupied', the HMM estimate of occupancy probability is about 0.58, while the corresponding

estimate of our method is around 0.91 at time step 3 in Fig. 4(c). The reasons for the difference in estimation results and the quick response to state transition that is observed in our method are revealed by the resulting transition models from training that are shown in Fig. 5. The transition model of the conventional HMM is fixed, as shown in Fig. 5(a), while the transition model of our method is adaptive to the values of other sequences as input to the IOHMM. For example, at time step 3 in Fig. 4(a), the neighbor sequence shows the value $'1'$ indicating of an 'occupancy event' for the sequence in concern (object is moving in), and our model captures this information by an increased transition probability from 'free' to 'occupied' as shown at time step 3 in Fig. 4(d) and by the transition model in Fig. 5(b). Similarly, our model responds to the $'1'$ in the neighbor sequence at time step 6 in Fig. 4(a), which indicates the sequence of concerned to become 'free' (object is moving out), with an increased transition probability from 'occupied' to 'free' as shown at time step 6 in Fig. 4(d) and by the transition model in Fig. 5(c). During periods without transitional events, our model shows almost no possibility of state transition as in Fig. 5(d), while the model based on the HMM still maintains same level of transition tendency as in Fig. 5(a).
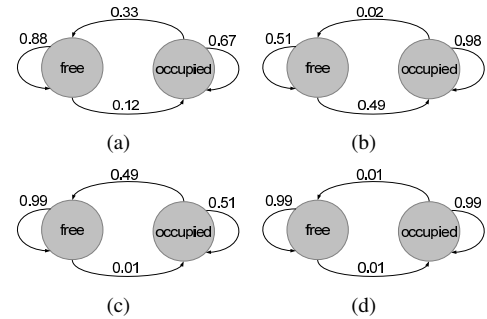

Fig. 5. Transition models from training. The homogeneous model from the conventional HMM is shown in (a). The inhomogeneous models from our method corresponding to time steps 3, 6 and 19 in Fig. 4(c) are shown in (b), (c) and (d) respectively.

The log-likelihood of all data from our training result is -8.578497, while that from the conventional HMM is -14.903643. This further proves that our method is more capable of explaining data that involve substantial dynamics and where the dynamics change over time.

IV. EXPERIMENTS

The goal of these experiments is to verify the capability of the method we propose to capture aspects of motion patterns of dynamic objects in the environment. To this end, we choose a coarse representation of the environment by setting the gird size for both dimensions as $0.3m$ in the occupancy grid map. Our experimental results show that this representation is sufficient for studying the motion pattern of people in our experimental environment - a corridor in an office building. It is worth noting that for motion with high agility, a fine representation with a small grid size should be used.

A. Experimental setup

We used a SICK LMS200 laser range-finder for collecting 2D data at approximately 37Hz. The environment chosen is one corridor of our center that is shown in Fig. 6. 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 approximately where the picture was taken and started data logging at 10:04am. The data collection process went on till 20:23pm, corresponding to 10 hours and 19 minutes of laser scans.
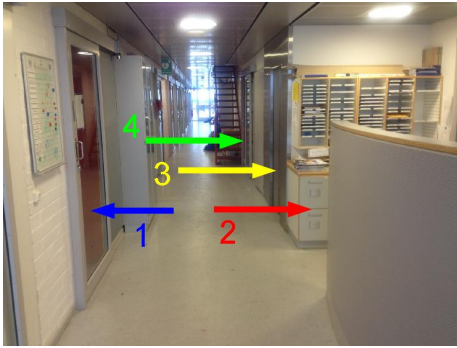

Fig. 6. Experimental environment.

B. Generating occupancy observations

The endpoint model [18] is applied to generate occupancy observations for each cell in the occupancy grid map. Cells along the laser ray are treated equally and assigned a 'free' observation, and the cell where the laser ray ends is assigned an 'occupied' observation. These observations form sequences for all cells in the map.

Our training method based on EM mainly learns how the transition probabilities depend on the input by maximizing the observation likelihood for the cell in concern. It is important to note that for this training the parameters to be learned in ρ are only concerned about the parts of the data that involve state transitions, which are corresponding to the 'occupancy events' such as those shown in Fig. 4(a). In the presence of such events, the EM based method responds to the transition occurrences and optimizes towards the best transition probabilities by assigning proper values for parameters in ρ. So we extract short segments from the data that contain 'occupancy events' for training and this significantly speeds up the training process, especially considering that interesting events happen in a comparatively occasional manner during the prolonged data collection period.

On the other hand, the training of methods based on the conventional HMM such as [8] requires using the data in full length, as they calculate one set of fixed transition probabilities corresponding to the average dynamics in the environment. We believe that this difference in training methodology is caused by the inhomogeneous nature of

our method in contrast to those based on the homogeneous model.

C. Training results

The macroscopic model from our training results that are achieved by using the procedure in Section III-C.5 in an offline manner with the collected data, which is overlaid with the floor plan of the building, is shown in Fig. 7. The arrows indicate tendency of objects moving from the cell where the arrow starts to the cell where it points to. The length of arrows is proportional to contribution of the starting cell as a neighbor input to the transition parameters ρ of the ending cell. In our current implementation, no training is carried out for cells on the edge of the occupancy grip map, as they have one or two neighbors missing, so there are no arrows pointing to them.

As shown by Fig. 7, most cells have longer arrows straight along the corridor than sideways ones (note the scales of the two axes are not equal), which is in accordance to the way people walk in a corridor. What can also be read from the figure is that sideways arrows do appear in significant magnitude in areas right outside the office, the elevator, the kitchen and the doorway to the bathroom (compare with Fig. 6).

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. Knowledge of this model enables generating online predictions about the future states of grid cells based on a few initial observations, and thus inferring about possible moving trajectories of objects in the environment. In this case, real observations along the trajectory may not be available, and estimates of the occupancy state can be used as input to drive the model.

In its current setting, the model we learned only captures the spatial aspect of the global dynamics in the environment, due to extracting occupancy events. But our method does capture temporal correlation locally. It is possible to add the temporal axis on the global scale by segmenting data, for example by the hour, and classifying results for each segment. The predictive power of our method is restricted by the first order model we use. It is also possible to extend this to higher orders and thus make use of, for example, velocity of people. These are targets for our future work.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed an IOHMM based method for modeling motion patterns of dynamic objects. In contrast to the conventional HMM, which is based on a homogeneous Markov chain, in our model the transition probabilities are conditional on the input (observations in neighbor cells at the previous time step) and is variant with time, resulting in an inhomogeneous Markov chain. Thus the dynamics of our model are adapted to the input. In the local area, the spatial-temporal correlation are captured. Globally, the collection of learned parameters for all cells provides an overview of the object moving trend in the entire environment.

When operating on the cell level, our model is 'event' driven. When there is an event happening in the neighborhood, this immediately drives the corresponding transition probability of changing state to a high value. On the other hand, in case of no event occurrences in the neighbor cells, the transition probabilities of changing state stay very low.

Compared with object tracking approaches, such as [19], we take a different point of view, which is whether a location is occupied by an object. Then we connect these occupancy events and collectively infer about motion pattens in the whole environment. Hence there is no need to explicitly track multiple individual objects, providing significant convenience in data processing such as associating data segments to different objects being tracked.

Our method is based on regularities in the way that objects move in human living environments. For example, people normally walk straight in the corridor and only turn sideways in cases like entering the kitchen or an office. It is obviously valuable to be able to capture these regular tracks and utilize the information contained in the 'habitual' behavior in various potentially beneficiary robotic tasks such as long-term mapping and/or localization with dynamic objects as well as navigation in human populated environments involving avoiding or interacting with people such as [9].

ACKNOWLEDGMENT

REFERENCES

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

[2] A. Walcott-Bryant, M. Kaess, H. Johannsson, and J. J. Leonard, "Dynamic pose graph SLAM: Long-term mapping in low dynamic environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 1871–1878.

[3] H. Moravec and A. Elfes, "High resolution maps from wide angle sonar," in *IEEE Int. Conf. Robotics and Automation (ICRA)*, March 1985, pp. 116–121.

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

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

[6] J. P. Saarinen, H. Andreasson, T. Stoyanov, and A. J. Lilienthal, "3D normal distributions transform occupancy maps: an efficient representation for mapping in dynamic environments," *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1627–1644, 2013.

[7] D. Arbuckle, A. Howard, and M. Matari, "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)*, 2004, pp. 409–414.

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

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

[10] O. Cappe, "Online EM algorithm for hidden Markov models," *Journal of Computational and Graphical Statistics*, vol. 20, no. 3, pp. 728–749, 2011.
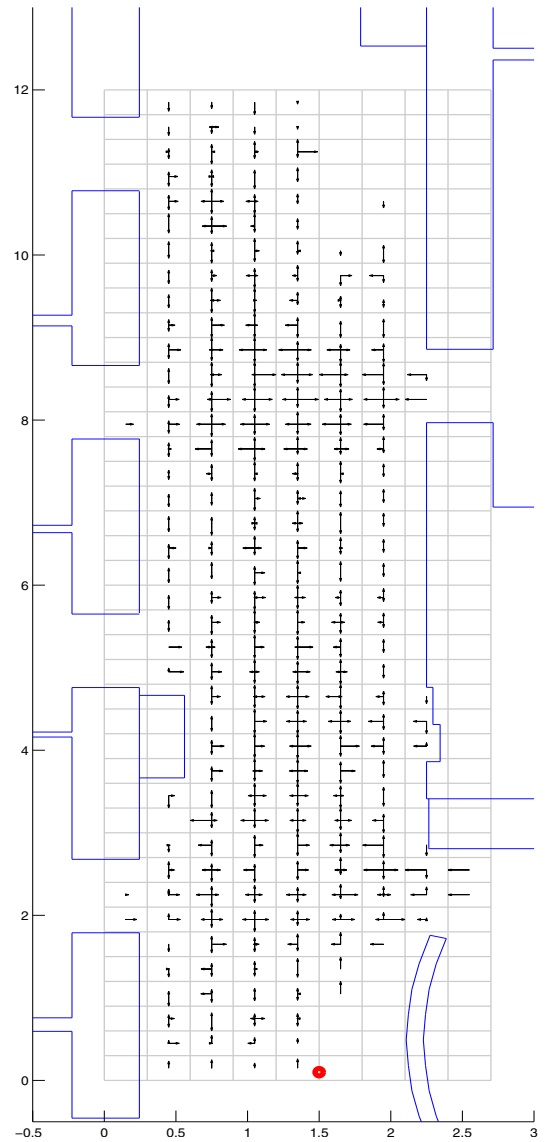
Fig. 7. Training results of the transition parameters ρ. The red dot at the bottom indicates the sensor location.

[11] Y. Bengio and P. Frasconi, "Input-output HMMs for sequence processing," *IEEE Trans. on Neural Networks*, vol. 7, no. 5, pp. 1231–1249, 1996.

[12] Y. Bengio, "Markovian models for sequential data," *Neural Comput. Surv.*, vol. 2, pp. 129–162, 1999.

[13] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*. Springer, 1995.

[14] Y. Li and H. Shum, "Learning dynamic audio-visual mapping with input-output hidden Markov models," *IEEE Trans. on Multimedia*, vol. 8, no. 3, pp. 542–549, 2006.

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

[16] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[17] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–285, 1989.

[18] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2006.

[19] M. Luber, G. D. Tipaldi, and K. O. Arras, "Place-dependent people tracking," *The International Journal of Robotics Research*, vol. 30, no. 3, pp. 280–293, 2011.