

Building a Human Behavior Map from Local Observations

Zhan Wang, Patric Jensfelt and John Folkesson

Abstract—This paper presents a novel method for classifying regions from human movements in service robots’ working environments. The entire space is segmented subject to the class type according to the functionality or affordance of each place which accommodates a typical human behavior. This is achieved based on a grid map in two steps. First a probabilistic model is developed to capture human movements for each grid cell by using a non-ergodic HMM. Then the learned transition probabilities corresponding to these movements are used to cluster all cells by using the K-means algorithm. The knowledge of typical human movements for each location, represented by the prototypes from K-means and summarized in a ‘behavior-based map’, enables a robot to adjust the strategy of interacting with people according to where they are located, and thus greatly enhances its capability to assist people. The performance of the proposed classification method is demonstrated by experimental results from 8 hours of data that are collected in a kitchen environment.

I. INTRODUCTION

Service robots must interact with people while collaborating with them. Inevitably the people and the robots will have to move through the same working environment. Therefore, the robots can benefit by mapping human motion patterns within the environment. For example, in a dining area, in which areas do people stand while working. These areas might be good ones for the robot to offer help. If there are doorways in which people move quickly, the robot should avoid getting in the way of them [1].

For service robots, the complete environment model should include the knowledge of local human behaviors. That is, in each specific location (for example, a cell in the grid map), what are the typical human activities. The behavior layer (also termed as the ‘behavior-based map’ in this paper) in the representation hierarchy of the environment is as important as the metric layer (such as where is the kitchen and its dimension) and the semantic layer (such as if the segmented object in a certain location is a dining table). Although it is different from the other two which are about static objects that can be perceived all the time, the behavior layer plays an important role in the robot’s operation in human-populated environments. It enhances the robot’s capability to serve people by choosing the best strategy according to the areas where they are located.

In the research of obtaining semantic information through human-environment interactions, the affordance of objects has been used. The concept of affordance [2] was initially introduced into the computer vision community to model

the relationship between the human behavior and the environment, including objects, events and places. To quote [2], “Affordance: A situation where an object’s sensory characteristics intuitively imply its functionality and use.” In [3] a spatial-affordance map is used to give a spatial dependence to the Poisson process parameters used for person tracking. Our behavior map is similar but not designed specifically for the data association during people tracking. Rather our map models a number of general motion patterns and is more intended for prediction than tracking.

Our work on region classification is motivated by the functionalities or affordances of places. For example a doorway is for people to pass without stopping, a hallway is for people to walk in all directions or to stop for chatting/looking, a display window area outside a shop is for people to wander and do ‘window shopping’, a dining area in the kitchen is for people to sit and eat, a corridor is for people to pass and occasionally stop. Although not directed exploited in the proposed method, these underlying functionalities or affordances consolidate the feasibility of grouping locations according to human behaviors.

Due to the dynamic nature of humans in motion, the knowledge of human activities in each location of the environment can only be learned from data collected over long periods. For a robot to learn motion patterns it is important that the model can be learned from local observations. The robot will not be able to follow people but will rather only observe partial trajectories. Our learning approach requires observations in a 9-cell patch of a grid map to learn the model of the center cell. This is an important point, that the robot must interact with people based on where they are. It is not always realistic to expect that the robot will have tracked the person to that location. This is why methods such as the IOHMM as in [4] or the SVM learning of whole track classes as in [5] were not chosen here.

Summarizing, we are not tracking people but rather learning about the space they travel through. The robot should be able to learn the behaviors associated with regions of the space in which it moves. This mapping will be useful to it in planning tasks such as cleaning, navigation or interacting with people. The robot is not able to follow people to obtain long complete tracks but rather short segments of tracks. The method of learning must work on such observations.

In this paper we provide an unsupervised learning approach. In our method, a probabilistic model is developed to capture people’s moving activities in public areas represented by a grid map. In the light of the successful application of HMMs on modeling motion patterns of dynamic objects [6] [4], we exploit the HMM to learn the human moving

Zhan Wang, Patric Jensfelt and John Folkesson are with Computer Vision and Active Perception Lab (CVAP), Centre for Autonomous Systems (CAS), CSC, The Royal Institute of Technology, KTH, Stockholm 100 44, Sweden {zhanw, patric, johnf}@kth.se

tendency in each individual grid cell, but in a novel manner that explicitly models the transition tendency from the current cell to a neighbor cell and the tendency of staying. Collectively, the transition probabilities in all cells represent the moving trend of humans in broader areas. Then based on these transition probabilities, the K-means algorithm is applied to cluster all cells into different categories representing places with different typical human activities, such as minimal movements and mostly staying, only passing in certain directions without staying and mostly passing with moderate staying, etc.

The contributions of this work include:

- 1) A probabilistic method is designed to model human moving tendency including staying in the current location. Compared with other region classification methods, which are mostly only based on the most dominating local activities [1] [5], our method provides a more detailed summary of local activities to the region classifier, which is then capable of dealing with more complicated situations. In contrast to trajectory-based motion learning algorithms [7], our model focuses on the local area and does not require tracking humans in motion, which greatly simplifies the implementation and application of our method. As illustrated by the experimental results, local human behaviors are sufficient for region classification and building the behavior-based map.

- 2) An unsupervised learning method based on the K-means algorithm is developed to classify regions with different human movements.

II. RELATED WORK

There are many works on modeling the behavior associated with a particular person such as [8]. In this paper however we are interested in modeling the behavior associated with a particular region of space.

The occupancy grid map [9] and related grid-based representations have been widely used in the study of dynamics of the environment. The dynamic occupancy grid [6], which utilizes a HMM [10] with a two-state Markov chain to model the occupancy of a cell, successfully relaxed 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 over a certain timescale. The online training procedure provides adaptive capabilities from the recent tendency. In [11], 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 are that each individual cell is modeled independently and that correlations between cells are completely ignored. In the Conditional Transition Maps developed by Kucner et al. [12], 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 [13], local spatial-temporal correlations in the dynamics of human movements is captured by using Input-Output HMMs. However, these methods do not model staying in the same cell. In our work, we explicitly characterize staying by the self-transition probability to the cell itself. This enables the model to describe the special case of movements, for example, when a person is standing in the corridor and talking with others or sitting in the dining area of the kitchen and having lunch.

In [14], a classification method is proposed to understand which locations are appropriate for a shopping-assistant robot to wait to serve customers. A grid-based representation with 33 cm square cells (similar to the cell size that we use, 40 cm) is used to estimate if the robot’s waiting in a cell will disturb pedestrians from counting the number of people that passed the cell per hour from collected data. In our method, the cells are classified by a more detailed description of human activities in a cell, and thus providing the robot with more informative guidance in interacting with humans.

There are several works on classifying cells by the human activities within. The work in [1] provides several techniques for motion prediction with the aim of enabling the robot to proactively approach customers, for example, in a shopping mall. Collected trajectories are categorized by using a SVM according to the local behaviors. Global behaviors are learned through clustering using the dial pulse (DP) matching method. However, the similarity measurement used is not sensitive to trajectories with no or little overlap. Each local partition (from non-uniform sized spatial division) of the environment is color-coded according to its dominant local motion primitive class, as well as the transitions between adjacent areas, which are computed for each pair of adjacent areas by counting the transitions. In [5], grid cells are categorized based on the most prevalent motion pattern learned from a pre-filtering SVM, and sub-patterns are learned by using the Partitioning Around Medoids (PAM) algorithm with an improved dissimilarity measurement. Both of these methods provide good results. In our method, we learn a probabilistic model to provide a full summary of local activities and input to the region classifier.

Human activities have been used to infer about semantics in the environment. In [15], human body movements, which are detected and recognized from data collected by motion sensors mounted on the human body, together with the location of the human are used to update the semantic map. Grabner et al. [16] have proposed an affordance detector which is defined as an actor-object matching problem. The functionality is handled as a cue complementary to appearance, rather than being considered after appearance-based detection.

III. MODELING HUMAN MOVEMENTS IN GRID MAPS

Similar to the occupancy grid map representation [9], we divide the overall map into grid cells in the two dimensional space. Accordingly, the original data sequence is discretized in the time domain such that the observed human moves only

to a neighbor cell or stays in the same cell in one time step. The dynamics of human movements is studied on the cell level.

The continuous movement of a human 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 multiple time steps subject to temporal and spatial discretization. It is worth noting that before reaching any of the neighbor cells the human may stay in the current cell for multiple time steps. Figure 1 shows the possible directions that a human in a cell may move.

We model the human moving tendency from the current cell to a neighbor cell by a non-ergodic HMM [17], and its structure facilitates capturing the spatial-temporal correlation between the two cells involved in a movement.

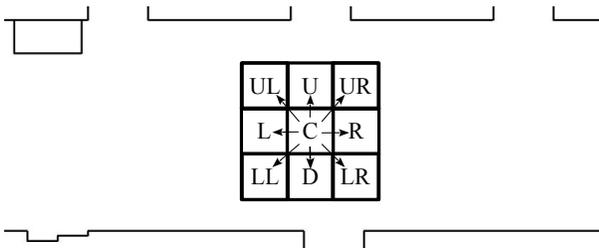


Fig. 1. Illustration of one-step movements in a corridor environment. The human is originally located in cell C and may move to any of the eight neighboring cells. Note that the human may stay in the cell C for multiple time steps before starting to move.

A. 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 nine states [17]. Different from normal left-to-right HMMs, our model has eight 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 as shown in Figure 1 is attached to each cell in which the current cell is denoted as C . The process of one-step movement for an individual cell starts from the human entering the current cell. In each of the following time steps, the human can either stay in the current cell or move in one of the eight directions (left, right, up, down, upper left, lower right, upper right, lower left) as illustrated in Figure 1, and the process ends when the human reaches any of the neighbor cells. The latent variable x_t represents the location of the human in the nine cells involved in the current one-step movement at time step t , and it can take nine states $\{C, L, R, U, D, UL, LR, UR, LL\}$, corresponding to the nine cells (center, left, right, up, down, upper left, lower right, upper right, lower left). Let z_t be a random variable that represents the observation of the human in a cell at time step t , and it can take nine states $\{C, L, R, U, D, UL, LR, UR, LL\}$.

The process under consideration is only one move from the current cell, and thus the states

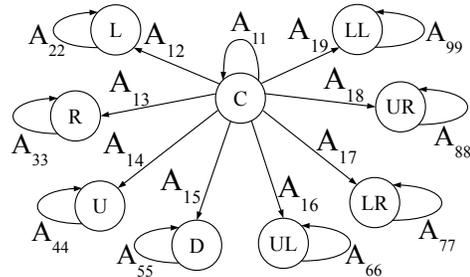


Fig. 2. Transition diagram of our model of one-step movement by a variant of the left-to-right HMM.

$\{L, R, U, D, UL, LR, UR, LL\}$ are specified as ‘absorbing or final states’. Let A_{lk} represent the state transition probability from the l^{th} state to the k^{th} state, where $l, k = 1, \dots, 9$. Figure 2 shows the transition diagram of our HMM-based model. The corresponding transition matrix is

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} & A_{17} & A_{18} & A_{19} \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (1)$$

The ones on the main diagonal indicate that the corresponding states $\{L, R, U, D, UL, LR, UR, LL\}$ are absorbing states.

B. Training Procedure

As the HMMs in different cells are independent conditioned on the observation set, the training for an individual HMM can be performed separately. We adapt the scheme of the generalized EM algorithm [18] 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. This is 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 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 $\pi = (1, 0, 0, 0, 0, 0, 0, 0, 0)$ and not re-estimated.
- At the beginning of the forward-backward scheme [19], set the transition probabilities $A_{lk} = 0, (l > k)$ and $A_{lk} = 0, (l < k, l > 1)$. In the transition matrix (1), the off-diagonal elements (except those in the first row) are zero because the transition from any of the

absorbing states in $\{L, R, U, D, UL, LR, UR, LL\}$ to other absorbing or non-absorbing states is not allowed. The value of these elements will remain zero during the whole training process. The self-transition probabilities of the absorbing states are set as one, $A_{lk} = 1, (l = k, l > 1)$.

- The probability $\beta(x_t)$ in the forward-backward scheme is defined as $\beta(x_t) = p(z_{t+1}, \dots, z_T | x_t)$, ($t = 1, \dots, T - 1$), where T is the total number of time steps in the current training sequence [19]. The initial condition $\beta(x_T)$ is set as

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

Suppose the training data for the current cell are a set of P short-observation sequences, $\{z_1^{T_p}(p); p = 1, \dots, P\}$. Each sequence $z_1(p), \dots, z_{T_p}(p)$ consists of all observations of the corresponding one-step movement which lasts for T_p time steps. For example, the sequence C, C, C, C, R corresponds to the one-step movement starting from the current cell, then staying for three time steps, and then moving to the cell on the right. 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 and the model parameters,

$$\begin{aligned} \gamma(x_t) &= p(x_t | z_1^T, \hat{\Theta}) \\ \xi(x_{t-1}, x_t) &= p(x_{t-1}, x_t | z_1^T, \hat{\Theta}) \end{aligned} \quad (3)$$

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

$$Q(\Theta, \hat{\Theta}) = \sum_{t=2}^T \sum_{j=1}^K \sum_{k=1}^K \xi(x_{t-1,j}, x_{t,k}) \ln A_{jk} + \sum_{t=1}^T \sum_{k=1}^K \gamma(x_{t,k}) \ln p(z_t | x_{t,k}), \quad (4)$$

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 9.

During the process of training the model parameters using the generalized EM algorithm, the sum of the expectation of the likelihood in equation (4) for all the P short-observation sequences of the current cell is maximized with respect to Θ , and therefore the estimated values of parameters in Θ are obtained.

IV. REGION CLASSIFICATION BY USING K-MEANS

Human living environments are structural and functional. Human behaviors in these environments are defined by the functionality or affordance of each local region, such as the hallway, the pathway, the dining area, the sink area, the fridge and cabinet area. For example, the most frequent behavior of people in a corridor is passing while where to turn is defined by the shape of the corridor. People are most likely to be

sitting on the seats (and eating) when they are in a dining area of a kitchen. Collectively, the local human motion in each small area, such as a grid cell, provides valuable information for the representation of the environment on the behavior level. The functionalities or affordances of different places allow the local areas in the environment to be classified by local human activities.

We have studied the human activities in Section III, where the learned transition probabilities of each grid cell explicitly model human moving tendencies in eight directions and staying. For each cell, we stack the nine transition probabilities into a vector

$$[A_{11}, A_{12}, \dots, A_{18}, A_{19}]. \quad (5)$$

Then we apply a linear re-scaling of the data points, which is known as standardizing. In the resulting data points each of the variables has zero mean and unit standard deviation.

The K-means algorithm is used to cluster the grid cells into different categories. In the experiments of Section V, the parameter K is set manually with examination of human behaviors in the environment. The human motion prototypes can be obtained from the resulting centroids after clustering using K-means. These prototypes represent typical human movements in a cell of each region category, such as mostly staying with moderate motion in the area next to the sink of the kitchen, as shown in the experimental results of Section V.

V. EXPERIMENTS

A. Experimental setup

The goal of these experiments is to verify the capability of the method we propose to capture the dynamics of local human movements and to classify regions based on these movements. Considering the size of a human body, we choose a coarse representation of the environment by setting the grid size as $0.4 \times 0.4m$ in our map. The experimental results show that this representation is sufficient for studying the human movement pattern in our experimental environment - a kitchen in an office building, as shown in Figure 3.

We used a SICK LMS200 laser range-finder for collecting 2D data at approximately 37Hz. The laser range-finder was set up on a table in the kitchen at the height of a human's waist and its location is shown as the star in Figure 3(a). The data collection process started from 11:12am on a weekday and went on for 8 hours and 6 minutes. The time is chosen to cover when there are most human activities in the kitchen during the day including the periods for lunch and afternoon tea. The model learned from this data represents averaged characteristics of human motion. Learning for each specific time period of the day and studying how the model evolves with time is among our future work.

B. Generating observations for training

We applied the endpoint model [20] to generate occupancy observations for each cell in the map. With temporal discretization, the original prolonged occupancy observations

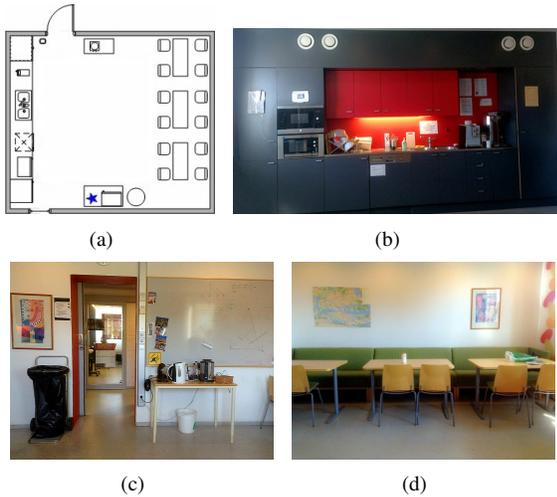


Fig. 3. The kitchen Environment. (a) Layout of furniture in the kitchen. The blue star at the bottom indicates the laser sensor. (b) Furniture on the left side of (a). From the left to right, they are the fridge, the microwave, the dish dryer, the sink and the cabinet for tea. (c) The door, the dustbin and the table with hot water cookers on the top side of (a). (d) The dining area on the right side of (a).

related to one cell are transformed into a sequence of interesting events of entering and exiting a grid cell by analyzing the occupancy in the cell of concern and its eight neighbor cells. As the entering/exiting events in different cells are treated separately, tracking of human across several cells is not necessary, which significantly simplifies the data processing.

As discussed in Section III-B, the training of our motion model requires short-observation sequences (starting from the current cell and ending in one of the neighbor cells, i.e. the absorbing states) in order to make full use of training data. For each cell, a set of short-observation sequences in the form of z_1, \dots, z_T are generated. For example, in the observation sequence C, C, C, R , the human is originally located in the current cell, stays for two time steps and then moves to the neighbor cell on the right.

C. Results of motion learning and region classification

Figure 4 shows the motion learning and region classification results in the grid map which is overlapped with the part of the floor plan corresponding to the kitchen. The two sets of results are shown together for the ease of understanding. In each cell, the transition probabilities from the current cell to the eight neighbor cells are indicated by the eight arrows, and the transition probability of staying in the current cell is indicated by the red line segment. The classification result is illustrated by the color of the solid square in each cell and cells in the same class have squares with the same color.

In Figure 4, each class of cells with the same color represents a typical human movement. For example, the brown cell category features exiting the current cell in all directions and all the cells are in the open space in the kitchen where people tend to move in all directions. The cyan cell category features staying in the cell and the most evident examples are cells in the dining area where people

sit and have lunch, around the table with hot water cookers and beside the fridge, the dish dryer and the sink.

There are some areas where people tend to come and stay for a while, as indicated by the red line segment. For example, people tend to walk to the sink and stay to wash their mug or lunch box then leave. In some areas, people tend to stop and talk with others. Other spots where people tend to stay for a short while include those close to the fridge where they store or fetch food and the hot water cooker table where they get boiled water. The corresponding cells in these areas have red line segments with significant length.

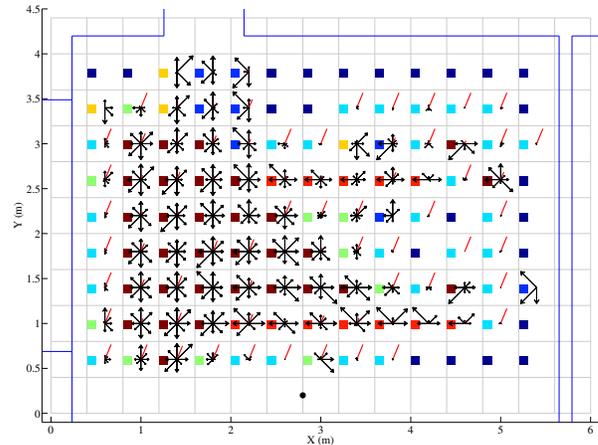


Fig. 4. Results of motion learning and region classification. The arrows and the red line segment indicate transition probabilities of moving to eight neighbor cells and staying in the current cell respectively. The colored square in each cell indicates the region class.

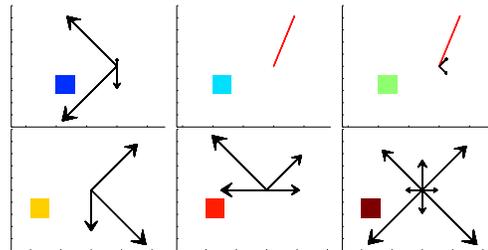


Fig. 5. Centroids from the K-means algorithm for each type of region. With the same color-coding as Figure 4 and Figure 6, the colored square represents the corresponding region category. The deep blue area is meaningless so the centroid is not included. For each centroid, the arrows and the red line segment indicate tendency of moving and staying respectively.

The representative activity of each type of cells can be seen from the centroids resulting from the K-means algorithm, as shown in Figure 5. For example, the one (with a brown square) on the right of the lower row in Figure 5 features moving to all directions, corresponding to the brown area in Figure 4. These centroids (prototypes) can be considered as local behavior primitives and the real data points are approximate combinations of them.

The clustered local regions consisting of cells of the same color constitute a segmentation of the environment by human movements and result in a behavior-based map. For example, in Figure 6, the brown area represents the open space in

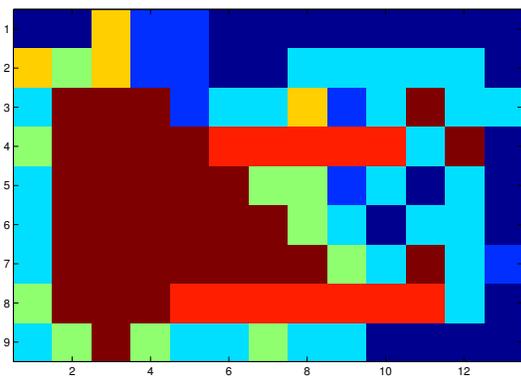


Fig. 6. Results of region classification. Each color represents a type of region with the same color-coding in Figure 4.

the central area of the kitchen where people tend to walk in all directions. The red areas indicate the two pathway from the brown area to the dining area where people tend to mostly walk straight with the constraint of the wall and the table of hot water cookers. The sky blue area at the door is the pathway from outside to the brown area where people tend to walk straight but not sideways with the constraints of the dustbin and the table of hot water cookers. In this area, people do not tend to stop to avoid getting into others' way of entering/exiting the kitchen. The deep blue areas correspond to regions where there are not enough training data in the cells because they are mostly corresponding to walls and hot water cookers on the table.

D. Computational cost

Currently the motion learning and region classification method is implemented in Matlab. The training for each cell is carried out separately, and then all cells are clustered by the resulting transition probabilities using the K-means algorithm. For the results that are shown in Figure 4, the total computational cost of training the 117 cells using the 9872 short-observations is 308 seconds and the time for clustering is 0.13 seconds on a laptop (2.90GHz Intel Core i7 processor, 8.00GB of RAM).

VI. DISCUSSIONS

The region classification method described in above sections results in a behavior-based map summarizing local human behaviors, as illustrated by Figures 4 and 6, which can be readily integrated into a planning component for a service robot. For example, the behavior-based map in Figure 6 can be combined with an occupancy grid map and utilized in a LPA* based path planner [21]. The local behaviors in each cell correspond to a value of cost or award in the objective function, depending on the goal, for example, meeting or avoiding the human. The behavior-based map can also be combined with other cues of the environment and generate decisions for the robot, such as where to be stationed to serve people. For example, stationary persons are likely to be in regions where the representative centroid has a long red line segment, which represents the tendency of staying in the cell,

and these regions should be prioritized in the task of serving coffee [22]. In the following, we describe more application scenarios.

In the proposed method, the clusters of cells resulting from K-means do not directly provide information about the meaning of each type of region. However, by analyzing the centroids, the robot can infer about the general motion pattern in the corresponding type of region, which can be exploited to make decisions on the interaction strategy. For example, the centroid with a sky blue square on the left of the upper row in Figure 5 (please distinguish this from the deep blue squares in Figures 4 and 6 which represent cells where there are not enough training data) indicates that cells of this category feature “moving in vertical directions without stopping”. Hence long narrow areas consisting of cells of this type (such as the doorway area colored as sky blue in Figures 4 and 6) are likely to be “fast passing areas” and the robot should not stay in these areas to avoid interrupting people’s routine walks.

If semantics of the environment is also available, then the robot is further facilitated to make decisions about its action. In the above example, if it is known that the object next to the sky blue area in the upper part of Figures 4 and 6 is a door, then the robot is able to pick the suitable place to station itself and wait for people to be served [14], i.e. next to the door but not in the “fast passing area”, the doorway.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed an unsupervised approach to region classification based on human movements for service robots with a grid map. For each cell, we model all possible human movements with transition probabilities by using the HMM. These transition probabilities are then used to classify all cells of the map into categories with different typical human behaviors, forming a behavior layer of the map.

Together with the metric and semantic layers, the behavior layer contributes to a more thorough representation of the environment for service robots and plays an important role in their interaction with humans. The classification of regions based on activities in each areas enables the robot to estimate the typical movement of the person from local observations and thereby adjust the interaction strategy.

In our method, local behavior primitives are learned and used for inferring the type of cells. This allows segmenting the environment into small areas by the typical human activity corresponding to the affordance of different type of places. In terms of summarizing long-term observations of human movements, our method takes a location-based point of view, which is different from the trajectory-based methods focusing on characterizing long continuous trajectories [7]. In this regard, our method has the practical advantage of working without tracking moving people, and thus does not need to deal with the multi-target tracking problem. Our future work includes implementing and running the proposed algorithm on our robot as well as investigating how human movements and region classes change over time during the day.

ACKNOWLEDGMENT

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

REFERENCES

- [1] T. Kanda, D. F. Glas, M. Shiomi, and N. Hagita, "Abstracting peoples trajectories for social robots to proactively approach customers," *IEEE Transactions on Robotics*, vol. 25, no. 6, pp. 1382 – 1396, 2009.
- [2] J. Gibson, *The Ecological Approach to Visual Perception*. Houghton Mifflin, 1979.
- [3] 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.
- [4] Z. Wang, R. Ambrus, P. Jensfelt, and J. Folkesson, "Modeling motion patterns of dynamic objects by IOHMM," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014.
- [5] S. Xiao, Z. Wang, and J. Folkesson, "Unsupervised robot learning to predict person motion," in *IEEE International Conference Robotics and Automation (ICRA)*, 2015, pp. 691–696.
- [6] D. Meyer-Delius, M. Beinhofer, and W. Burgard, "Occupancy grid models for robot mapping in changing environments," in *AAAI conference on artificial intelligence*, 2012.
- [7] B. T. Morris and M. M. Trivedi, "A survey of vision-based trajectory learning and analysis for surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1114–1127, 2008.
- [8] T. Ikeda, Y. Chigodo, D. Rea, F. Zanlungo, M. Shiomi, and T. Kanda, "Modeling and prediction of pedestrian behavior based on the sub-goal concept," *Robotics*, p. 137, 2013.
- [9] H. Moravec and A. Elfes, "High resolution maps from wide angle sonar," in *IEEE Int. Conf. Robotics and Automation (ICRA)*, 1985, pp. 116–121.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] Z. Wang, P. Jensfelt, and J. Folkesson, "Modeling spatial-temporal dynamics of human movements for predicting future trajectories," in *AAAI Workshop: Knowledge, Skill, and Behavior Transfer in Autonomous Robots*, 2015, pp. 42–48.
- [14] T. Kitade, S. Satake, T. Kanda, and M. Imai, "Understanding suitable locations for waiting," in *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2013, pp. 57 – 64.
- [15] W. Sheng, J. Du, Q. Cheng, G. Li, C. Zhu, M. Liu, and G. Xu, "Robot semantic mapping through human activity recognition: A wearable sensing and computing approach," *Robotics and Autonomous Systems*, vol. 68, pp. 47–58, 2015.
- [16] H. Grabner, J. Gall, and L. J. V. Gool, "What makes a chair a chair?" in *Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1529–1536.
- [17] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, "An introduction to the application of the theory of probabilistic functions of a markov process to automatic speech recognition," *Bell System Technical Journal*, vol. 62, no. 4, pp. 1035 – 1074, 1983.
- [18] 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.
- [19] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [20] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2006.
- [21] S. Koenig, M. Likhachev, and D. Furcy, "Lifelong planning a*," *Artificial Intelligence*, vol. 155, no. 1-2, pp. 93–146, 2004.
- [22] G. D. Tipaldi and K. O. Arras, "I want my coffee hot! learning to find people under spatio-temporal constraints," in *IEEE International Conference Robotics and Automation (ICRA)*, 2011, pp. 1217–1222.