Robust Localization of Persons Based on Learned Motion Patterns

Grzegorz Cielniak[†‡] **Maren Bennewitz**[†] **Wolfram Burgard**[†]

[†]Department of Computer Science, University of Freiburg, 79110 Freiburg, Germany
[‡]Department of Technology, Örebro University, 70182 Örebro, Sweden

Abstract. Whenever people move through their environments they do not move randomly. Instead, they usually follow specific trajectories or motion patterns corresponding to their intentions. Knowledge about such patterns may enable a mobile robot to robustly keep track of persons in its environment. This paper proposes a technique to derive a Hidden Markov Model (HMM) from learned motion patterns of people. This HMM is able to estimate the current and future positions of persons given knowledge about their intentions. Experimental results obtained with a mobile robot using laser and vision data collected in a typical office building with several persons illustrate the reliability and robustness of the approach. We also demonstrate that our model provides better estimates than an HMM directly learned from the data.

1 Introduction

Whenever mobile robots are designed to operate in populated environments, they need to be able to perceive the people in their neighborhood and to adapt their behavior according to the activities of the people. Knowledge about typical motion behaviors of persons can be used in several ways to improve the behavior of a robot since it may provide better estimates about current positions of persons as well as allow better prediction of future locations. In this paper we present an approach for estimating future poses of persons based on motion patterns learned from data. Our system derives a Hidden Markov Model (HMM) from learned motion patterns and uses a combination of laser and vision data to update the HMM. Based on this information our system can reliably estimate the locations of multiple persons in the environment.

Recently, a variety of service robots has been developed that have been designed to operate in populated environments. These robots for example, have been deployed in hospitals [6], museums [4], office buildings [1], and department stores [5] where they perform various services e.g., deliver, educate, entertain [13] or assist people [12, 8, 11]. A variety of techniques has been developed that allows a robot to track people in its vicinity [14, 7]. Additionally, several authors have used models of people's motions to adapt the behavior of a mobile platform according to predicted movements [17, 16, 9]. These approaches however can only predict short-term motions. Bui et al. proposed an Abstract Hidden Markov Model (AHMM) to predict people's motion [3]. In contrast to our approach, they do not take into account that different persons might have individual motion patterns. Using our technique, however, we are able to distinguish between persons and incorporate this into the estimation process. The technique described in this paper is an extension of the approach recently proposed by Bennewitz et al. [2]. We describe how to derive an HMM from the previously learned motion patterns. This Hidden Markov Model allows the robot to maintain a belief about the current location of a person. Furthermore, we present an extension which is able to estimate the positions of multiple moving persons in the vicinity of the robot.

The paper is organized as follows. In the next section we explain how we generate Hidden Markov Models to predict motions of persons. Section 3 describes our laser- and vision-based approach to keep track of people. In Section 4 we present several experiments illustrating the robustness of our approach for estimating the positions of single and multiple persons using laser and vision data with a mobile robot. We also give results indicating that our models provide better estimates than Hidden Markov Models directly learned from the observations.

2 Deriving Hidden Markov Models from Learned Intentions

People usually do not permanently move. Rather they typically move between so-called staying areas or resting places. Our approach therefore assumes that the motion patterns of persons are given in the form of M trajectories θ_m each consisting of a sequence of points interconnected by line segments. Obviously, the initial point and the final point of each motion pattern must correspond to resting places. To derive a Hidden Markov Model from the given typical trajectories of a person we therefore distinguish two types of nodes. The first class are the initial and final nodes that correspond to the resting places which are the start and end positions of the given trajectories. To connect these nodes we introduce so-called intermediate nodes which lie on the motion patterns. In our current system we use a sequence of L_m intermediate nodes $\nu_m^1, \ldots, \nu_m^{L_m}$ for each intention θ_m. The intermediate nodes are distributed over θ_m such that the distance between two consecutive nodes is $\Delta_\nu = 50$cm. Given this equidistant distribution of the intermediate-nodes and assuming a constant speed v with standard deviation σ_v of the person, the transition probabilities of this HMM depend on the length Δ_t of the time interval between consecutive updates of the HMM as well as on v and σ_v. In our current system, this value is set to $\Delta_t = 0.5$secs. Accordingly, we compute the probability that the person will be in node ν'_m given it currently is in ν_m and given that the time Δ_t has elapsed as:

$$p(\nu'_m \mid \nu_m, \Delta_t) = \int_{\nu'_m - \frac{\Delta_\nu}{2}}^{\nu'_m + \frac{\Delta_\nu}{2}} \mathcal{N}(\nu_m + v \cdot \Delta_t, \sigma_v, x)\, dx. \tag{1}$$

Here $\mathcal{N}(\nu_m + v \cdot \Delta_t, \sigma_v, x)$ is the value of the Gaussian with mean $\nu_m + v \cdot \Delta_t$ and standard deviation σ_v at position x. We currently define the same transition probabilities for all intermediate nodes and assume that the persons are moving at constant speed. Of course one could specify different transition probabilities for individual nodes depending on the actual velocity. The transition probabilities for the resting places are computed partially based on a statistics about the average time period which elapses before the person starts to move after staying at the corresponding resting place. Furthermore, we observed the persons and counted how often a person started to move on a particular trajectory. Thereby, we learned the information needed to specify the transition probabilities for the resting places.

Please note that the resulting model can be regarded as a two-level Abstract Hidden Markov Model [3]. Whereas the higher-level goals of this AHMM correspond to the resting places of the person, the lower-level goals are the nodes along the paths to the high-level goals.

3 People Tracking and Identification

In principle one would have to maintain a belief over the joint state space of all persons in order to correctly approximate the posterior about the locations of all persons in the environment. This approach, however, is usually not feasible since the complexity of the state estimation problem grows exponentially in the number of persons or dimensions of the state space. Therefore, we approximate the posterior by factorizing the belief over the joint state space and by considering independent beliefs over the states of all persons. To maintain the individual beliefs we need to be able to update the HMMs for the persons based on observations made by the robot. To keep track of persons and to identify them our systems combines laser and vision information. To identify persons in the laser data our system extracts features which are local minima in the range scans that come from the legs of the persons. Additionally, it considers changes in consecutive scans to more reliably identify moving people. Furthermore, we use a map of the environment and also extract features not belonging to static objects. Since several persons can be in a range scan, we apply a set of Kalman filters. In each Kalman filter, the state x of a person is represented by a vector $[x, y, \delta x, \delta y]'$. Whereas x and y represent the position of the person, the terms δx and δy represent the velocity of the person in x- and y-direction. Accordingly, the prediction is carried out by the equation:

$$x_{r+1}^- = \begin{vmatrix} 1 & 0 & t_r & 0 \\ 0 & 1 & 0 & t_r \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} x_r \tag{2}$$

where t_r is the time elapsed between the measurement z_{r+1} and z_r. Usually, the laser range sensors only provide information about the position of a person. Since the velocities δx and δy are also part of our state space but cannot be observed directly, the measurement matrix projects onto the first two components of the state space. Accordingly, the predicted measurement at step $r+1$ is:

$$z_{r+1}^- = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{vmatrix} x_{r+1}^-. \tag{3}$$

To track multiple persons in the range scans, we apply independent Kalman filters, one for each feature. To solve the data association problem, we apply a nearest neighbor approach, i.e. we update a filter using the observation z_{r+1} that is closest to z_{r+1}^-. New filters are introduced for observations from which all predictions are too far away. Furthermore, filters are removed if no corresponding feature can be found for one second.

We also need to be able to identify a person in order to appropriately update the belief about the location of that person. To achieve this we additionally employ the vision system of our robot. To identify a person, we proceed as follows: Every time the laser-based people tracker detects a person in the field of view of the camera, an image is collected and following three steps are applied:

1. *Segmentation:* The size of a rectangular area of the image containing the person is determined.
2. *Feature extraction:* We compute a color histogram for the area selected in the previous step.
3. *Database matching:* To determine the likelihood of a particular person, we compare the histogram computed in step 2 to all prototypes existing in the database.

To determine the area in the image corresponding to a feature detected by the laser tracking system, we rely on an accurate calibration between the camera and the laser and we use a perspective projection to map the 3D position of the person in world coordinates to 2D image coordinates. Whereas color histograms are robust with respect to translation, rotation, scale and to any kind of geometric distortions they are sensitive to varying lighting conditions. To handle this problem we consider the HSV (Hue-Saturation-Value) color space. In this color model the intensity factor can be separated so that its influence is reduced. In our current system we simply ignore this factor. Throughout all our experiments we could not find any evidence that this factor negatively affected the performance of the system. The image database is created beforehand.

To compare a given query histogram I with a prototype M in the database we use the normalized intersection norm $H(I, M)$ [15]. This quantity can be computed as:

$$H(I,M) = \frac{\sum_{j=1}^{n} \min(I_j, M_j)}{\sum_{j=1}^{n} M_j}, \tag{4}$$

where I and M are color histograms both having n bins. One advantage of this norm is that it also allows to compare partial views, i.e. when the person is close to the camera and only a part of it is visible.

Thus, to incorporate an observation sequence $z_1, \ldots, z_R$ into the HMM of a person we apply the following recursive Bayesian update scheme:

$$p(\nu \mid z_1, \ldots, z_R) = \alpha \cdot \begin{cases} p(z_R \mid \nu) \cdot p(\nu \mid z_1, \ldots, z_{R-1}) & \text{if } z_r \text{ is only a range measurement} \\ p(z_R \mid \nu) \cdot H(I,M) \cdot p(\nu \mid z_1, \ldots, z_{R-1}) & \text{if } z_r \text{ is a range and vision measurement} \end{cases} \tag{5}$$

Whenever the robot receives only laser range information, we update all HMMs using this information, since we do not know which person actually was perceived by the robot. If, however, the measurement includes range and vision information, we update each HMM using the range information but also proportional to the likelihood that this particular measurement was reflected by the person corresponding to the particular HMM. In our current system, the likelihood $p(z_r \mid \nu)$ of an observation z_r given the state ν is computed using a Gaussian distribution which depends on both, the variance in the current estimate of the tracking system and the variance σ used during the learning of the intentions.

4 Experimental Results

The technique described above has been implemented and evaluated using data acquired in an unmodified office environment. The experiments described in this section are designed to illustrate that our approach can be used to robustly estimate the positions of persons. Additionally, we compare the performance of our algorithm to that of standard HMMs directly learned from the data.

To analyze the applicability of the HMMs for the prediction of the locations of a person we performed several experiments with our robot Albert (see Figure 1) in our office environment. The map of the environment as well as the structure of the HMM used for the experiments is depicted in the left image of Figure 2. During most of the experiments the robot was moving along the corridor of our department with a speed of up to 40cm/sec. Thereby its task was to maintain a belief about the positions of the persons.

4.1 Using the Derived HMM to Keep Track of a Single Person

The first experiment is designed to illustrate that our approach is able to reliably estimate the position of a person in its environment. In this experiment, a single person was moving in our department (see Figure 1) and the task of the robot, which itself did not move, was to estimate the positions of this person. Especially, we were interested in the probability that the person stayed at the correct resting place.

The second left image of Figure 2 plots for different resting places the probability that the person stays at this particular place. Whereas the x-axis represents the individual time steps, the y-axis indicates the probability. The graph also includes the ground truth, which is indicated by the corresponding horizontal line-pattern at the .9 level. As can be seen from the figure, the system can reliably determine the current position of the person. During this experiment it predicted the correct place of the person in 93% of the time.

Fig. 1. Robot Albert tracking Maren while she is walking through the environment.

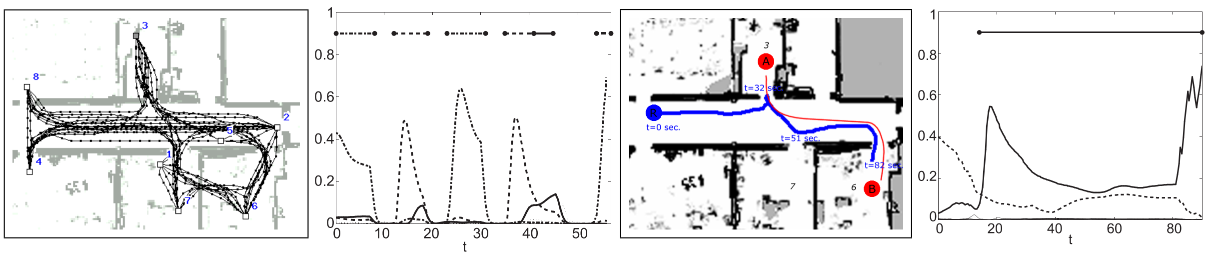

Fig. 2. Hidden Markov Model derived from learned intentions (left image) and evolution of the probability for a person of being at the different resting places over the time and the ground truth (second left). In the experiment depicted in the third image the person walked from A to B (thin line). The robot was moving down the corridor (thick line) thereby looking into rooms at the indicated time-steps. The right image again shows the evolution of the probability of the different resting places over the time.

In the next experiment the robot was programmed to drive along the corridor and to inspect the different rooms in order to detect and locate persons. The path of the robot (thick line) as well as the person's path (thin line) are depicted in the third image of Figure 2. The robot's path is labeled with the time-steps at which the robot looked into the corresponding rooms. The right image of Figure 2 plots for different resting places the probability that the person stayed at this particular place (the solid line corresponds to room 6, the dotted line to room 3, and the thin line to room 7). As before, the graph also includes the ground truth. Please note, that our robot only has one laser-range scanner which means that it can only cover the 180 degrees in front of it. This is for example why the probability that the person stayed at the resting place 3 (dotted curve) increased again at around time step 40 after the robot had inspected the corresponding room. There was some probability "flowing back" to that resting place. The robot could not sense that no person was moving behind it while it was looking into the room which contains resting place 7.

4.2 Estimating the Locations of Multiple Persons

The image database which was created beforehand contains for each person one histogram that was built from 20 images. As explained in Section 3, to incorporate the similarity measure provided by the vision system into the HMM of a person, we multiply the likelihoods provided by the laser tracking system with the similarity measure if the current estimate of the laser tracker is in the field of view of the camera. Otherwise, we simply update the HMMs of all persons. As an application example consider the situation depicted in the left image of Figure 3. In this particular situation two persons (Cyrill and Greg) were walking along the corridor within the perceptual field of the robot. The right image of Figure 3 shows the estimate of the laser-based people tracking system at the same point in time. The corresponding image obtained with the robot's camera is shown in the left image of Figure 4. Also shown there are the two segments of the image that correspond to the two persons detected with the laser. Figure 5 shows the resulting histograms of the two persons. The center image of Figure 4 plots the similarities of the histograms of the two segments to the individual prototypes stored in the data base. Finally, the right image of Figure 4 depicts the HMM for Cyrill (who is the left person in Figure 3). As can be seen, the probabilities indicated by the size of the rectangles are higher for the states that correspond to the true location. Throughout this experiment the robot was able to predict the correct location of the persons in 79% of all cases.

4.3 A Comparison to Standard HMMs

The second experiment is designed to demonstrate that an HMM that takes into account the people's intentions allows a better prediction than a standard HMM that is directly generated from the observed trajectories of the persons and

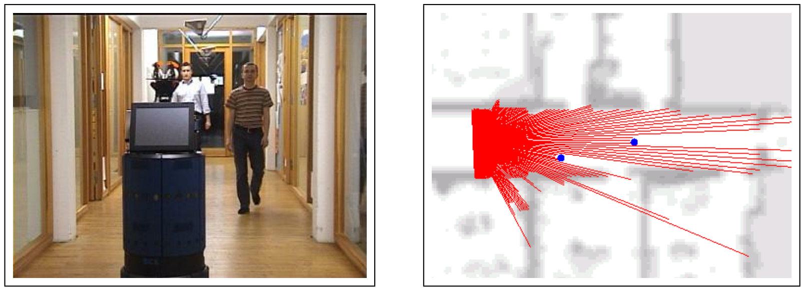

Fig. 3. Typical scene with two persons walking along the corridor (left image) and corresponding estimate of the laser-based people tracking system (right image).

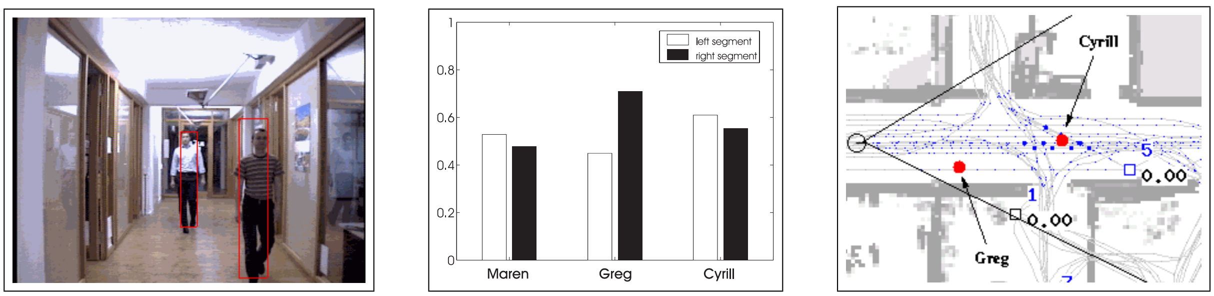

Fig. 4. Segmentation of the two persons from the image grabbed with the camera of the robot (left image), similarity of the histograms (see Figure 5) of the segments to the data base prototypes (center image), and resulting posterior after incorporating the two segments shown in the left image into the belief over Cyrill's position (right image)).

that does not take into a account the clustered trajectories. To evaluate the performance of the two different approaches we chose two motion patterns from those depicted in the left image of Figure 2. The first pattern is the one leading from resting place 7 via the office containing resting place 6 to the resting place 2. The second one is the motion pattern between the places 6 and 5. We defined a standard HMM over the possible states of the person in the $\langle x, y, dx, dy \rangle$ space where x and y were discretized in 15cm patches; dx and dy encode 9 possible incremental moves per cell. The transition probabilities were learned from the trajectories corresponding to both motion patterns by counting. We randomly chose a position along the trajectories of both patterns as the observed position of the person. The states of the HMM were initialized according to the observation model (see Section 3). After convergence of the HMM we measured the likelihood of the final destination. We compared this value to those obtained by the HMM generated by our algorithm for the trajectories corresponding to these two intentions. We repeated this experiment for different locations along the trajectories of both patterns and determined the average probability of the true goal location. Whereas we obtained an average of .74 with our model, the corresponding value of the standard HMM is .56. This illustrates that our model leads to better results and that the standard independence assumption of HMMs is generally not justified in this application domain. Please note that similar observations have been reported by Murphy [10]. In contrast to a standard HMM our model automatically chooses the transitions that correspond to the actual intention of the person.

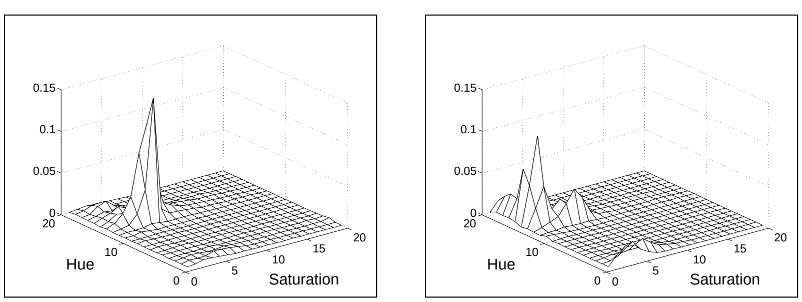

Fig. 5. Corresponding histograms for the two segments shown in the left image of Figure 4.

5 Conclusions

In this paper we presented a method for automatically deriving an HMM from typical motion patterns of persons. We also presented techniques to update the resulting HMMs using laser range data and vision information. We then demonstrated that we can use our HMMs to estimate the positions of persons in the environment of a mobile robot.

Our approach has been implemented and applied successfully to data recorded in a typical office environment. In practical experiments we demonstrated that our method is able to use learned motion patterns to reliably predict states of multiple persons. The experiments have been carried out using a mobile robot equipped with a laser-range sensor and a vision system. We furthermore presented experiments indicating that standard HMMs directly learned from the same input data are less predictive than our models.

References

1. H. Asoh, S. Hayamizu, I. Hara, Y. Motomura, S. Akaho, and T. Matsui. Socially embedded learning of office-conversant robot Jijo-2. In *Proc. of the Int. Joint Conference on Artificial Intelligence (IJCAI)*, 1997.
2. M. Bennewitz, W. Burgard, and S. Thrun. Using EM to learn motion behaviors of persons with mobile robots. In *Proc. of the Int. Conference on Intelligent Robots and Systems (IROS)*, 2002.
3. H. Bui, S. Venkatesh, and G. West. Tracking and surveillance in wide-area spatial environments using the Abstract Hidden Markov Model. *Intl. J. of Pattern Rec. and AI*, 2001.
4. W. Burgard, A.B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114(1-2), 1999.
5. H. Endres, W. Feiten, and G. Lawitzky. Field test of a navigation system: Autonomous cleaning in supermarkets. In *Proc. of the Int. Conference on Robotics & Automation (ICRA)*, 1998.
6. S. King and C. Weiman. Helpmate autonomous mobile robot navigation system. In *Proc. of the SPIE Conference on Mobile Robots*, pages 190–198, Boston, MA, November 1990. Volume 2352.
7. B. Kluge, C. Köhler, and E. Prassler. Fast and robust tracking of multiple moving objects with a laser range finder. In *Proc. of the IEEE Int. Conference on Robotics & Automation (ICRA)*, 2001.
8. G. Lacey and K. Dawson-Howe. The application of robotics to a mobility aid for the elderly blind. *Journal of Robotics and Autonomous Systems (RAS)*, 23:245–252, 1998.
9. S. M. Lavalle, H. H. Gonzalez-Banos, G. Becker, and J.-C. Latombe. Motion strategies for maintaining visibility of a moving target. In *Proc. of the IEEE Int. Conference on Robotics & Automation (ICRA)*, 1997.
10. K. P. Murphy. *Dynamic Bayesian Networks: Representation, Inference, and Learning*. PhD thesis, University of California, Berkely, 2002.
11. N. Roy, G. Baltus, D. Fox, F. Gemperle, J. Goetz, T. Hirsch, D. Magaritis, M. Montemerlo, J. Pineau, Schulte J., and S. Thrun. Towards personal service robots for the elderly. In *Proc. of the Workshop on Interactive Robotics and Entertainment*, 2000.
12. C. Schaeffer and T. May. Care-o-bot - a system for assisting elderly or disabled persons in home environments. In *Assistive technology on the threshold of the new millenium*. IOS Press, Amsterdam, 1999.
13. R.D. Schraft and G. Schmierer. *Service Robots*. Springer Verlag, 1998.
14. D. Schulz, W. Burgard, D. Fox, and A.B. Cremers. Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. In *Proc. of the Int. Conference on Robotics & Automation (ICRA)*, 2001.
15. M. Swain and D. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1), 1991.
16. S. Tadokoro, M. Hayashi, Y. Manabe, Y. Nakami, and T. Takamori. On motion planning of mobile robots which coexist and cooperate with human. In *Proc. of the Int. Conference on Intelligent Robots and Systems (IROS)*, 1995.
17. Q. Zhu. Hidden Markov model for dynamic obstacle avoidance of mobile robot navigation. *IEEE Transactions on Robotics and Automation*, 7(3), 1991.

This article was processed using
the TeX macro package with ECMR2003 style