Multimodal People Tracking and Trajectory Prediction based on Learned Generalized Motion Patterns

Martin Weser, Daniel Westhoff, Markus Hüser, Jianwei Zhang
Institute Technical Aspects of Multimodal Systems
Dept. of Informatics, University of Hamburg
Hamburg, Germany
{1weser,westhoff,hueser,zhang}@informatik.uni-hamburg.de

Abstract— **A sensor-based model of a service robot's environment is a prerequisite for interaction. Such a model should contain the positions of the robot's interaction partners. Additionally to the actual positions of the partners it is important for the service robot to predict their possible future positions. This knowledge could for example be used to realize efficient path planning for delivery tasks. In this paper we propose an extensible framework for systems, that combine different sensor modalities in a general tracking system. Furthermore, human trajectories are predicted by deducing them from learned motion patterns. Exemplarily, a tracking system is implemented that fuses tracking algorithms in laser range scans as well as in camera images by a particle filter. The observed trajectories are generalized to trajectory patterns by a novel method which uses self organizing maps. Those patterns are used to predict trajectories of the currently observed persons. Practical experiments show that multimodality increases the system's robustness to incorrect measurements of single sensors. It is also demonstrated that a self organizing map is suitable for learning and generalizing trajectories. Convenient predictions of future trajectories are presented which are deduced from these generalizations.**

I. Introduction

Service robots will support everyday work in business or home environments more and more in the near future. Possible services are delivery tasks, cleaning services or home care. Recent developments in mobile robotics emphasize this [1], [2], [3]. But, the distribution and therewith further development of mobile robots is mainly dependent on the acceptance of the society. An important criteria for this acceptance is the robot's ability to interact with the environment.

Interaction is only possible if the robot has knowledge about the locations of its interaction partners. In general, this knowledge can only be generated using sensory input. An explicit specification of a dynamic environment is usually impossible.

The generation of such an environment model is a nontrivial task since the sensor calibration and measurements are susceptible to errors. In addition, the observed objects may be occluded and therefore not detectable by some sensors. We argue that this uncertainty about the environment decreases when different sensor-modalities with diverse qualities are used. The advantages of the different sensors can complement one another using appropriate fusing methods. In the following we present a scalable robust system, that fuses arbitrary multimodal data containing trajectories of moving objects.

Besides the interaction partner's actual position, it is another important ability for a service robot to predict people's future positions. For example, the robot could intercept people for delivery tasks. Predictions about motions of dynamic objects are based on a motion model, which can be specified explicitly or learned from previous observed motions. We developed a novel approach to predict people's motion based on a motion pattern learned by a self organizing map.

The reminder of this paper is organized as follows: In section II an overview of existing research on trajectory prediction is given. Section III introduces the proposed framework for multimodal tracking. The laser- and camera-based tracking algorithms used in our implementation of the framework are presented. Preliminary experiments show the usefulness of the framework. The results of the experiments are used in section IV to explain the generalization of motion patterns and the experience-based prediction. Section V gives a conclusion.

II. Related Research

The research related to the proposed system for people tracking and trajectory prediction cuts into three areas: motion prediction, people tracking and filtering and fusion of multimodal tracking data.

A. Motion Prediction

Predicting the motion of objects is a commonly used to avoid collisions in path planning tasks. There are numerous approaches present in the literature to predict obstacle motion in this context. Until the last few years motion prediction was synonymous with the prediction of the objects position at the imminent following time step. This short-term prediction was mainly approached by modeling the object motion by a statistical process [4], [5], [6]. In [7] a Kalman filter is used to predict immediate following positions.

Even though several of this approaches might give accurate predictions, there actual benefit in relation to collision avoidance is questionable [8]. Short term predictions are also not sufficient to be used by higher level task planning processes. To predict object motions in a longer period of time, it is assumed that the motions are following determinated motion patterns.

Based on this idea of observable motion patterns, [9] clusters similar trajectories to trajectory prototypes by an expectation maximization algorithm. Partially observed trajectories are compared with these prototypes, to predict future motions of objects. It is assumed, that the object is moving on the prototype trajectory which is the most similar to the partially observed trajectory. Other methods [10], [11] use comparable techniques. The main difference lies in the method which is used to generate the trajectory prototypes. In [10] a pairwise clustering process is used. In [11] the target of the observed object's motion is estimated and the predicted path is generated by a path planning algorithm. The major drawback of that kind of systems is the impossibility to predict unusual trajectories.

B. People Tracking

The basis of the above mentioned prediction techniques is an accurate motion model, which rests upon previous observed trajectories. To ensure the quality of the observations it is already common to fuse data of multiple sensor modalities. Two frequently used sensors for tracking applications are cameras and laser range finders.

1) Camera Tracking: If static cameras are used, background subtraction is a common technique to separate foreground objects. In [12] every pixel is assigned with a statistical color probability of the observed background by a mixture of gaussians. This probability is used to determine the pixel's background membership in each frame. In [13] and [14] an occurrence model is used to track certain objects. This model contains the color appearance and a probability mask, which represents the probability of each pixel to be part of the object. Another approach to track certain objects is presented in [15]: Here a color histogram represents the model. This approach is applied in our system and it is described in section III in detail. Since this approach uses no background subtraction, it is suitable to track people in images recorded by mobile cameras .

2) Laser Tracking: In [16], [17] and [18] background subtraction is used with laser range scans for object tracking. The systems differ in the way they generate the background. They all have in common that the background is modeled as a probability density function over the range measurements. For laser range finders mounted on a mobile robot it is not suitable to use such background models since the background measurements are changing permanently. Therefore, in [19] human legs are registered only via their size and shape.

C. Filtering and Fusion of Multimodal Tracking Data

Both the tracking algorithms in laser range scanners and in camera images mostly include filtering of the position estimates. This is reasonable since all measurements contain errors. The filtering becomes more important if several sensors are used. The integration and weighting of the different sensors is mostly included in the filtering algorithm. In [19] a Kalman filter is used to integrate tracking algorithms in laser range scans and camera images. A particle filter is presented in [20] which fuses audio and video information. [21] applies another approach which deals with multimodality. The different tracks gained by different sensor modalities are connected via an anchoring method.

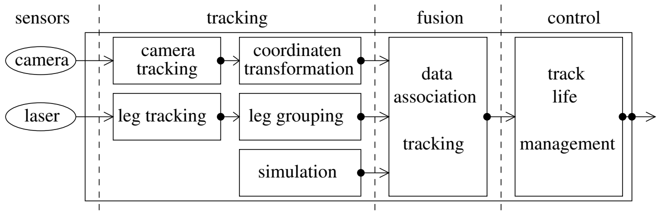

Fig. 1. Implemented components of the tracking system. The black dots symbolize the common interface described in section III-A.

III. PEOPLE TRACKING AND SENSOR FUSION

Both, camera tracking as well as laser tracking, have their own specific advantages and drawbacks. To build a robust and accurate tracking system, it is necessary to integrate several independent tracking algorithms working on the different sensors. With an appropriate fusion algorithm the specific advantages of the sensors could complement one another to decrease the overall error.

In the following, we introduce a conceptual framework, that is able to deal with an arbitrary number of sensors. Subsequent, we describe how a particle filter algorithm is applied to fuse and filter data of different sensor modalities. At last, the tracking algorithms for the measurements of laser range scanners as well as for camera images are presented.

A. Conceptual Framework for Multimodal Tracking

A typical multiple target tracking system consists of four blocks: sensor hardware, sensor processing and single sensor tracking, track fusion and association and track life management. A tracking system should be modular to allow the addition, removal and exchange of sensors and sensor processing algorithms. Therefore, the most important aspect of a tracking system becomes its ability to filter and fuse the results from individual sensors.

We developed a framework, that contains the above mentioned blocks. The schematic block diagram for this framework is shown in figure 1. Our implementation of the data association block includes filtering and data fusion using a particle filter. To ensure the extensibility of our system, we defined a general interface, that has to be implemented by the modules for low-level sensor processing as well as higher-level modules. Through this interface an arbitrary number of tracks is provided for the following module where each track consists of

- the current position and velocity of the tracked object,
- the uncertainty about the position and velocity and
- an unique identity number.

The structure of this framework is mainly motivated by software technical encapsulation, substitutability of algorithms and extensibility concerning further sensor modalities.

This structure provides efficient development, maintenance and testing capabilities.

B. Sensor Fusion and Filtering

We consider the problem of tracking as the detection of a state of a target. Therefore, we model the state x_t of a tracked person at time t as a four-dimensional vector $(x, y, \delta x, \delta y)^T$. This vector describes not only the position on the ground plain but also the velocity of the person. Since measurements of sensors contain errors it is impossible to derive the actual state of observed persons in a non-probabilistic way. Generally, a probability density function (*pdf*) is used to represent the state. Nonlinear Bayesian filtering can be applied to determine this *pdf* taking every previous measurement into account. The Bayesian solution to derive a belief about the current state is a recursive discrete time approach. Since optimal solutions for nonlinear Baysian tracking can only be applied when certain constraints hold we used the particle filter in our implementation. The particle filter is an approximate nonlinear Baysian filter.

The particle filter is used in two stages of our tracking system. First, it is used in the tracking algorithms applied on the data of a single sensor. Second, it is used for the fusion of tracking results made on different sensors. Other nonlinear filters like the extended Kalman filter can replace the individual particle filters in our system since each module is encapsulated and does not demand a particular algorithm.

The basis of the particle filter is the *importance sampling*. A multi dimensional function $g(x)$ is factorized into two functions $g(x) = f(x)\pi(x)$, where $\pi(x)$ is interpreted as a probability density function with $\pi(x) \geq 0$ and $\int \pi(x)\,dx = 1$.

If a set of samples $\{x^i | i = 1, \ldots, i = N\}$ with $N \gg 1$ and distributed according to $\pi(x)$ is generated, the integral of the function $g(x)$ can numerically be approximated as

$$\int g(x)\,dx \approx \frac{1}{N}\sum_{i=1}^{N} f(x^i) \tag{1}$$

When the density function $\pi(x)$ is unknown, $g(x)$ will be approximated by a function $\tau(x)$ similar to $\pi(x)$ regarding that $\forall_x : \pi(x) > 0 \Rightarrow \tau(x) > 0$. Equation (1) reformulates to

$$\int g(x)\,dx = \int f(x)\frac{\pi(x)}{\tau(x)}\tau(x)\,dx \approx \frac{1}{N}\sum_{i=1}^{N} f(x^i)\omega(x^i) \tag{2}$$

with

$$\omega(x^i) = \frac{\pi(x^i)}{\tau(x^i)} \tag{3}$$

when the sample set x^i is distributed according to $\tau(x)$

Importance sampling is applicable to nonlinear estimation if a set of samples $\{x_t^i, \omega_t^i\}$ is chosen at each time t where $\omega_t^i \triangleq \omega(x_t^i)$ describes a weighting of the samples and $\sum_{i=1}^{N}\omega_k^i = 1$. If $f(x) = 1$ this yields

$$p(x_k|Z_k) \approx \frac{1}{N}\sum_{i=1}^{N}\omega_k^i \delta(x_k - x_k^i). \tag{4}$$

The weights ω_k^i follow from

$$\omega_k^i \triangleq \omega(x_k^i) = \frac{p(x_k^i|Z_k)}{\tau(x_k^i|Z_k)}, \tag{5}$$

For each new measurement the *pdf* describing the state of a tracked person is approximated by displacing each sample according to a probability function $\tau(x_k^i|x_{k-1}^i, Z_k)$. It is incidental that $\tau(x_k^i|Z_k) = \tau(x_{k-1}^i|Z_{k-1})\tau(x_k^i|x_{k-1}^i, Z_k)$. The approximation defined recursive as

$$\begin{aligned} p(x_k^i|Z_k) &= \frac{p(z_k|x_k^i, Z_{k-1})p(x_k^i|Z_{k-1})}{p(z_k|Z_{k-1})} \\ &= \frac{p(z_k|x_k^i)p(x_k^i|x_{k-1}^i)p(x_{k-1}^i|Z_{k-1})}{p(z_k|Z_{k-1})} \end{aligned} \tag{6}$$

The weightings ω_k^i follow by insertion of (6) in (5).

$$\begin{aligned} \omega_k^i &= \frac{p(z_k|x_k^i)p(x_k^i|x_{k-1}^i)p(x_{k-1}^i|Z_{k-1})}{p(z_k|Z_{k-1})\tau(x_{k-1}^i|Z_{k-1})\tau(x_k^i|x_{k-1}^i, Z_k)} \\ &= \omega_{k-1}^i \frac{p(z_k|x_k^i)p(x_k^i|x_{k-1}^i)}{p(z_k|Z_{k-1})\tau(x_k^i|x_{k-1}^i, Z_k)} \end{aligned} \tag{7}$$

The normalizing factor $p(z_k|Z_{k-1})$ is constant and can be precalculated for each measurement.

The use of measurements gathered on multiple sensors causes different measurement models. This affects $p(z_k|x_k^i)$ and $p(z_k|Z_{k-1})$ for each sensor.

See [22] and [23] for more detailed descriptions on Baysian filtering.

C. Laser-based Tracking

As presented in section II tracking algorithms which use laser range finders are often divided into two steps. First, they generate a background model and, second, they determine the measurement's background membership. The background model is often represented by a histogram over the range measurements at each angle. It is assumed, that the maximum of the histogram is caused by the background. This model is usually calculated in advance. The range measurements similar to the background distance are classified as background and discarded. This has the following drawbacks:

- The prior generation of the background model is time-consuming.
- If there are foreground objects present during the calculation of the background they are included in the background.
- These systems are unable to handle alterations of the background.

We developed a novel method to calculate the background's ranges which is updated with each measurement. The background distance $h_i(t)$ at time t and angle i is given by the following recursive equation:

$$h_i(t) = h_i(t-1) + \begin{cases} \varepsilon_1 & \text{für } h_i(t-1) < m_i(t) \\ (-\varepsilon_2) & \text{sonst} \end{cases} \tag{8}$$

The values of the increments ε_1 and ε_2 determine the adaptivity of the background model.

After background measurements are removed groups of foreground measurements are tracked with a particle filter.

D. Camera-based Tracking

An overview of common methods for camera-based tracking is given in [24]. In our system we use the approach presented in [15] since it is appropriate to be used with cameras mounted on a mobile robot.

People tracked in the camera image are represented by a weighted color histogram. Pixel are weighted with a monotone decreasing kernel function $K : \mathbb{R}2 \rightarrow \mathbb{R}$ which assigns smaller weights to the pixels which are farther from the center of a detected person. If the size of a person is denoted by $2h^*$, the probability of the object's color u can be calculated as follows:

$$\hat{q}_u = C \sum_{x \in X^*} K\left(\frac{x}{h^*}\right) \delta\left(b(x) - u\right), \tag{9}$$

where C denotes a normalization constant. The function $b(x)$ assigns the pixel to an index of the histogram's color bin. Therefore, a person located at the coordinate y in the image plane is represented by a color histogram:

$$\hat{p}_u(y) = C_h \sum_{x \in X_h^y} K\left(\frac{y-x}{h}\right) \delta\left(b(x) - u\right) \tag{10}$$

where h is the size of the target candidate. As a measure of similarity between two color histograms we chose the *Bhattacharyya coefficient*.

The goal of the tracking algorithm is to find the location y with the highest similarity between the color histogram of a person and a candidate located at y. This is achieved using the sample mean shift method described in [15].

E. Experiemtal Results

For our experiments we used two SICK laser range finders mounted on a mobile service robot and a camera stationary mounted in the laboratory of the TAMS institute. Due to the uncertainty of the camera tracking, which is caused by noisy measurements and changing illumination conditions we weighted the outcome of the laser tracking higher. In a first experiment we used both sensor modalities to increase the accuracy and robustness of the tracking algorithm. Figure 2 shows the results of the multimodal tracking.

In a second experiment the system observed 43 trajectories during a two hours period of time. These trajectories are shown in figure 3. They will be used in the following section to explain the experience-based prediction with real-world examples.

IV. Generalization and Prediction of Trajectories

In recent years experience based long term prediction of trajectories became popular [9], [10], [11]. The systems have in common that motions are generalized first to predict future trajectories subsequent. We propose another approach to generalize trajectories completely unsupervised using a *Self Organizing Map* (SOM).

The following section describes the basics of the SOM algorithm and extensions needed to learn trajectory patterns. Afterward, a prediction algorithm is presented, which is based on the learned trajectory patterns.

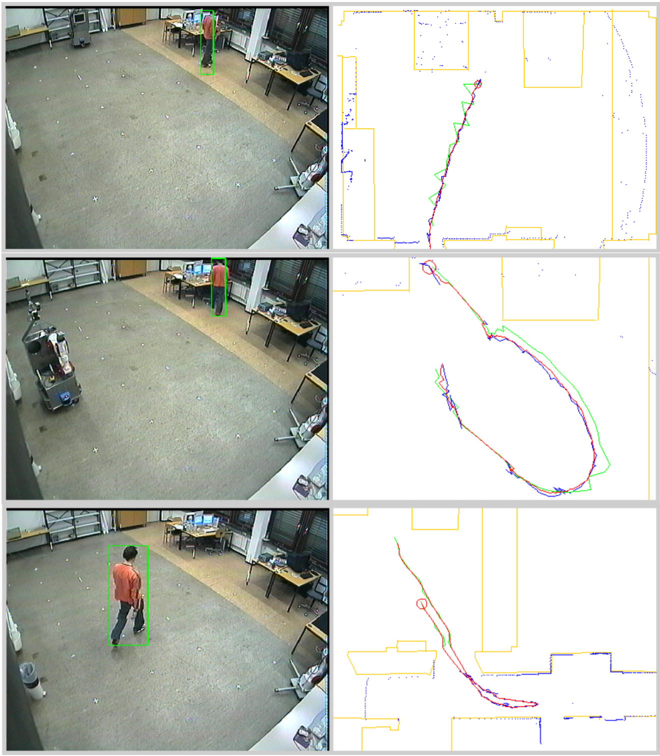

Fig. 2. Comparison of the sensor modalities: camera tracking (green) and laser tracking (blue) are fused by a particle filter (red). The greater variance of the camera tracking is obvious. The additional blue points are the current measurements of the laser range finders.

A. Learning of Trajectories using a Self Organizing Map

A SOM is a kind of artificial neuronal net. It is usualy used to map statistical data from a high-dimensional input space onto a set of reference vectors of a usually lower-dimensional topological space. A major feature of a SOM is topology conservation with respect to the neighborhood of the input set. That means that similar input vectors (IV) are mapped to the same reference vector (RV).

In our system the SOM is used to learn the observed trajectories of people. The topology conservation is directly used to generalize these trajectories to motion patterns. A subsequent clustering of trajectories is omitted. But, it is necessary to make the motion patterns which are inherent in the SOM explicit to be usable for the prediction. The following description of the SOM follows largely the descriptions of [25] and [26].

A SOM consists of a set of RV sometimes referred to as *nodes*. The RV are ordered in a topological space. In general, the alignment is arbitrary. We chose a quadratic structure to permit simple visualization and efficient storage in a two-dimensional array. The set of RV is depicted by $M = \{m^{ij} | m^{ij} \in \mathbb{R}^d\}_{i=1\ldots m_x, j=1\ldots m_y}$, where m_x is the width and m_y the height of the map. The dimension d of the RV and IV is equivalent.

During the learning phase the SOM is iteratively trained with the IV x^t. The RV $m^{cd} \in M$ most similar to x^t is referred

Fig. 3. The 43 trajectories observed during a two hours experiment.

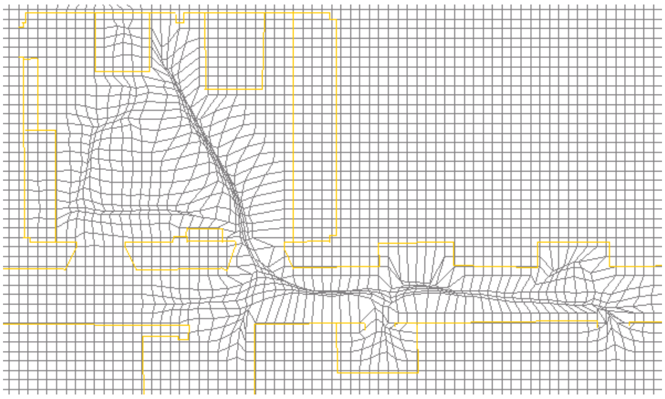

Fig. 4. This figure shows a converged SOM trained with real tracking results. The black lines represent the topological connections between the RV. The yellow lines represent a part of the map of the office environment of the TAMS institute.

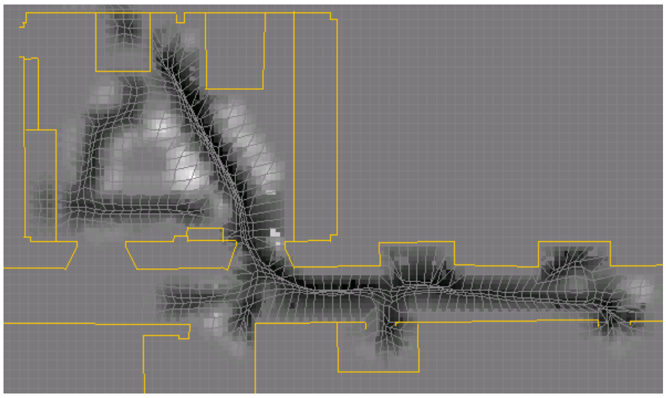

Fig. 5. The distribution of the MAAR values calculated on the RV of figure 4. The smaller the MAAR value the darker the color.

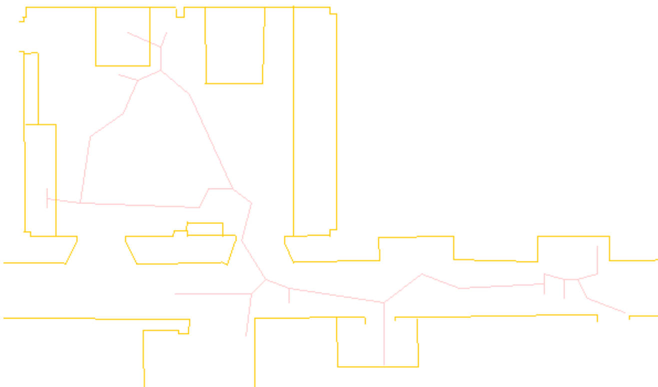

Fig. 6. The topology extracted from the MAAR distribution of figure 5 using Canny edge detection.

to as response node. As the degree of similarity we chose the euclidean distance in the input space. The response node and nodes in its topological neighborhood adept to the IV. We chose the Manhattan distance as the distance function for the topological space.

The learning step for each node $m^{ij}(t)$ at the time t is defined by the following learning rule:

$$m^{ij}(t+1) = m^{ij}(t) + n_t(d)l(t)\left(x^t - m^{ij}(t)\right),$$

where $l(t)$ is the learning rate and $n_t(d)$ is a neighborhood function depending on the Manhattan distance between m^{ij} and the response node m^{cd}.

The neighborhood function is a kernel function which specifies a smoothing factor. To enable local learning it is necessary that $n_t(d) \to 0$ for all $d > k_t$, where k_t is the size of the neighborhood. For the convergence of the SOM it is necessary that $l(t) \to 0$ or $k_t \to 0$ for $t \to \infty$. Figure 4 shows a converged SOM trained with real tracking results.

B. Extraction of learned Motion Patterns

After the SOM has converged the RV approximate the input data, i.e. the observed trajectories. It is self-evident to use a density estimation of the RV in the input space to characterize frequently used paths. Since the SOM algorithm does not generate new nodes, the distribution of the nodes after the learning process is based on a local re-allocation. Hence, next to every local density maximum there is a local minimum. That means transferred to the application of people tracking, that it is unlikely that a person stands beside a frequently used path. Since this does not agree with reality, we developed a novel method to characterize frequently used paths based on a SOM. Our method assigns to every node a value corresponding to the size of the minimum adjacent area which is spanned the four adjacent nodes and bi-linearly interpolated. This value is referred to as *MAAR value*. The smaller the MAAR value the more frequent a person was observed in this area. Figure 5 shows the MAAR distribution to the SOM of figure 4.

To extract the motion patterns from the MAAR distribution we used the gradient based Canny edge detector. The detected edges are transfered into a graph representation to permit the prediction described in the following subsection. The generalized motion pattern for the MAAR distribution of figure 5 is shown in figure 6

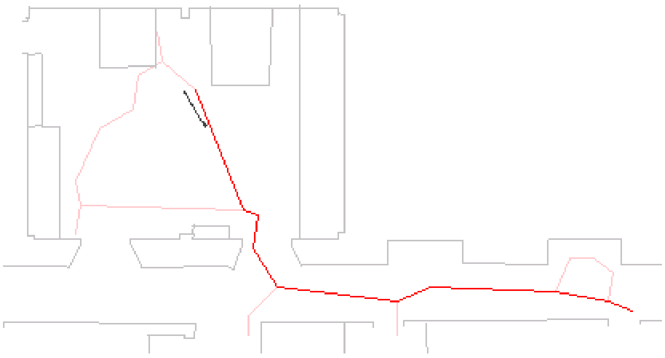

Fig. 7. The prediction (red) of the trajectory of an observed person (black) based on the generalized motion graph (pink).

C. Experiance based Trajectory Prediction

The graph extracted from the MAAR distribution represents frequently used trajectories in a generalized form. The trajectory prediction based on this motion patterns assumes that tracked persons normally move along the graph segments.

1) Mapping Graph Segments to Observed Persons: To decide which motion pattern matches best to an observed person we required a measure, which defines a correlation between a person and graph segment. We use the euclidian distance between a persons'position and the graph segments. The direction of the observed persons'motion is not considered for the mapping but is used later to determine the person's direction on the mapped motion pattern.

2) Prediction: Since experiments showed that the mapping is suitable, the motion pattern i.e. the graph segment which the observed person is moving along is known. With this, the long term prediction is to determine the sequence of path segments the person will be following subsequently. It is assumed that the general behavior of the tracked people does not change compared to the learning phase i.e. they will move close to the generalized trajectories.

To predict a person's motion, the probabilities for turns at graph branchings are learned. The long term prediction of a trajectory adds adjacent path segments to the mapped segment with respect to the probability of a turn at a graph branching. The prediction terminates if no subsequent path segment is available or the overall probability i. e. the product of the probabilities for each prior turn comes under a certain threshold. Figure 7 examplarily shows a prediction of the trajectory of an observed person.

V. Conclusion

In this paper we presented a framework where different sensor modalities can be integrated to build a multimodal tracking system. We implemented this framework for people tracking algorithms processing laser range measurements and camera images. The individual tracking algorithms for the different sensors as well as the fusion module use particle filters. Due to the modular design other filter algorithms can be used. With some preliminary experiments we showed that the robustness of the tracking can be increased when camera- and laser-based tracking is combined.

Further we presented a new approach to generalize observed trajectories to motion patterns using a self organizing map in order to predict future trajectories. We tested our generalization algorithm with more than 40 trajectories collected during an two hours observation session.

References

[1] R. D. Schraft, J. Neugebauer, and C. Schaefferand T. May. *Care-O-bot: Ein technisches Hilfssystem für unterstützungs- und pflegebedürftige Personen im häuslichen Bereich*, pages 234–244. Springer, 1998.

[2] S. Thrun, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Hähnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz. MINERVA: A second generation mobile tour-guide robot. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 1999.

[3] N. Roy, G. Baltus, D. Fox, F. Gemperle, J. Goetz, T. Hirsch, D. Magaritis, M. Montemerlo, J. Pineau, J. Schulte, and S. Thrun. Towards personal service robots for the elderly. In *Proceedings of the Workshop on Interactive Robotics and Entertainment (WIRE)*, Pittsburgh, 2000.

[4] Y. S. Nam, B. H. Lee, and M. S. Kim. View time based moving obstacle avoidance using stochastic prediction of obstacle motion. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 1081–1086, Minneapolis, 1996.

[5] S. Tadokoro, M. Hayashi, Y. Manabe, Y. Nakami, and T. Takamori. On motion planning of mobile robots which coexist and cooperatewith human. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, pages 518–523 vol.2, August 1995.

[6] N. H. C. Yung and C. Ye. Avoidance of moving obstacles through behavior fusion and motion prediction. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pages 3424–3429, San Diego, 1998.

[7] R. Madhavan and C. Schlennov. Moving object prediction for off-road autonomous navigation. In *Proceedings of the SPIE Aerosense Conference*, 2003.

[8] Amalia Foka. *Predictive Autonomous Robot Navigation*. PhD thesis, University of Crete, May 2005.

[9] Maren Bennewitz. *Mobile Robot Navigation in Dynamic Environments*. PhD thesis, University of Freiburg, Department of Computer Science, 2004.

[10] A. D. Vasquez and Th. Fraichard. Motion prediction for moving objects: a statistical approach. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3931–3936, New Orleans, LA (US), April 2004.

[11] A. Bruce and G. Gordon. Better motion prediction for people-tracking. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2004.

[12] W. E. L. Grimson, C. Stauffer, R. Romano, and L. Lee. Using adaptive tracking to classify and monitor activities in a site. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1998.

[13] A. Senior, A. Hampapur, Y-L Tian, L. Brown, S. Pankanti, and R. Bolle. Appearance models for occlusion handling. In *Second International Workshop on Performance Evaluation of Tracking and Surveillance Systems (ECCV)*, 2001.

[14] A. W. Senior. Tracking with probabilistic appearance models. In *International Workshop on Performance Evaluation of Tracking and Surveillance Systems (ECCV)*, pages 48–55, June 2002.

[15] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *Proceedings of the IEEE Conference on Computer Vision andÂ Pattern Recognition*, Hilton Head Island, South Carolina, 2000.

[16] H. Zhao, R. Shibasaki, and N. Ishihara. Pedestrian tracking using single-row laser range scanners. In *Proceedings of Computers on Urban Planning and Urban Management*, 2003.

[17] Ajo Fod, Andrew Howard, and Maja J. Mataric. Laser-based people tracking. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2002.

[18] H. Zhao and R. Shibasaki. Pedestrian tracking using multiple laser range scanners. In *Proceedings of Computers on Urban Planning and Urban Management*, Sendai, Japan, May 2003.

[19] S. Feyrer and A. Zell. Robust real-time pursuit of persons with a mobile robot using multisensor fusion. In *Proceedings of the 6th International Conference on Intelligent Autonomous Systems (IAS-6)*, 2000.

[20] H. Asoh, F. Asano, T. Yoshimura, K. Yamamoto, Y. Motomura, N. Ichimura, I. Hara, and J.ire Ogata. An application of a particle filter to bayesian multiple sound source tracking with audio and video information fusion. In *International Conference on Information Fusion*, pages 805–812, 2004.

[21] M. Kleinehagenbrock, S. Lang, J. Fritsch, F. Lömker, G. A. Fink, and G. Sagerer. Person tracking with a mobile robot based on multi-modal anchoring. In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication (ROMAN)*, pages 423–429, 2002.

[22] B. Ristic, S.Arulampalam, and N. Gordon. *Beyond the Kalman Filter*. Artech House, 2004.

[23] Yaakov Bar-Shalom, X.-Rong Li, and Thiagalingam Kirubarajan. *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc, 2001.

[24] James Russel. *Detecting Humans in Video Footage using Multiple Classifiers*. PhD thesis, University of Western Australia, 2004.

[25] T. Kohonen. *Self-Organizing Maps*. Springer- Verlag Berlin Heidelberg, 1995.

[26] L. Jain U. Seiffert. *Self-Organizing Neural Networkes*. Physica Verlag Heidelberg New York, 2002.