Track-Person Association Using a First-Order Probabilistic Model

Thomas Geier, Susanne Biundo
Institute of Artificial Intelligence
Ulm University
Ulm, Germany
thomas.geier@uni-ulm.de, susanne.biundo@uni-ulm.de

Stephan Reuter, Klaus Dietmayer
Institute of Measurement, Control and Microtechnology
Ulm University
Ulm, Germany
stephan.reuter@uni-ulm.de, klaus.dietmayer@uni-ulm.de

***Abstract*—This work addresses the problem of track association in person tracking. We propose a probabilistic model, based on Markov Logic Networks, that aims at associating the individual tracks emerging from a person tracking algorithm to the correct persons. For this purpose the continuous estimates of the object positions acquired by the tracking algorithm are mapped into discrete spatial regions, which are based on a floor plan of the environment. Experiments show that the described model is able to exploit the additional information contained inside the provided floor plan, and deliver good results compared to a state of the art person tracking algorithm despite the lossy discretization step. We discuss the engineered model in detail and give an empirical evaluation using an indoor setting.**

I. Introduction

This work is concerned with the problem of providing background-knowledge to the specialized application of person tracking using a first-order probabilistic model. We demonstrate that a hand-crafted model, built using Markov Logic Networks (MLN) [1], can help in solving the data association problem in tracking during situations where high occlusion prevents the correct association between past and new tracks. By leveraging additional information in form of a floor plan, we can resolve otherwise opaque situations.

In object tracking, a major problem lies in the task of data association — telling which measurements emerge from which object. Since this cannot be resolved perfectly, tracking algorithms produce errors like track confusions, where the same track ID is assigned to multiple real objects, or interruptions of tracks, where the same object spawns multiple, consecutive track IDs. This work is mainly concerned at resolving the second problem by reassociating track IDs with each other that originated from the same physical object.

The usage of a first-order probabilistic model like MLNs allows for an easier modeling task, because dependencies are represented by weighted first-order logical formulas instead of, e.g., conditional probability tables for the case of directed models like Bayesian Networks. In addition, the model is formulated in a lifted form and can be instantiated for the desired number of concurrent tracks or persons within the scene, which is not possible when using completely propositional models or specialized template models like dynamic Bayesian networks [2]. A model given in a lifted representation also makes it possible to leverage structure information contained within the lifted formulation for more efficient inference [3]; although this approach is not investigated here.

We have pursued the presented work with the idea of improving tracking results. But from a different point of view, it also describes how uncertain measurements from a laser range finder can be incorporated into a probabilistic knowledge-base via a preprocessing step using an object tracking algorithm. Extending this idea, the provided model represents only one module within a larger technical system that could be extended with additional information sources, like audio, video, or user input. As an application of the localization information, a technical system could display information differently depending on whether a user is within a certain range. Such information can be extracted directly from the proposed model in form of marginal probabilities or most probable explanations, depending on whether probabilistic uncertainty shall be carried on or not, respectively.

We evaluate our work in the context of an indoor situation, where multiple persons move in a two-room office, containing the laser range finder and several areas of interest like a printer or a coffee maker. We measure the quality of our model by the extent to which it is able to correctly reassociate object tracks emerging from the tracking algorithm with the correct persons inside the scene. We compare the quality of track association for six different hand-crafted floor plans to the number of track losses and confusions produced by a state of the art multi-object tracking algorithm.

The rest of the paper is laid out as follows. After we discuss related work, we give an overview of the applied tracking algorithm and introduce the concept of Markov Logic Networks. Then, we describe the investigated problem in detail and discuss the used MLN model. We give an empirical evaluation of the described setup and conclude with some possible extensions to the model and an overall discussion.

II. Related Work

In multi-object tracking, state dependent detection probabilities of objects are disregarded in most applications. Thus, a track disappears shortly after entering an occluded area and a new track is created when the object leaves the occluded area again. Consequently, one object is represented by different track IDs. Especially in scenarios where persons interact several times with a system, changed track IDs lead to the loss of the objects history. The multi-object Bayes filter [4] allows to integrate state dependent detection probabilities even if the scenario is characterized by a high object density [5]. In case of short term occlusions, the usage of state dependent detection probabilities leads to an improved track continuity. A direct integration of goals into the prediction of a persons' state is crucial, since the persons' action may be contradictory to the assigned goal.

Markov Logic Networks have been used by Sadilek and Kautz [6] for multi-agent activity recognition based on GPS data in a game of capture the flag. While their work can leverage more expert knowledge (the rules of the game), they do not encounter the data association problem present in the tracking scenario, since each person was carrying a personal GPS receiver. Tran and Davis [7] apply Markov Logic Networks to a parking lot surveillance scene using video data to recognize which person enters which car. They also track pedestrians across a scene and face the problem of data association. Sensory information emerges from image data and their focus lies in integrating different information sources that are all extracted from the same video stream. Track association was not inferred by the MLN but calculated (e.g. by color matching). Markov Logic Networks are also used by Singla and Domingos [8] for entity resolution in text mining. This is the problem of inferring which references refer to the same entity and it is similar to the data association problem in tracking. The two latter works use an `equals` predicate for identity maintenance, whereas we approach the problem using an association mapping to underlying entities. When grounding the model our approach only creates associations between currently instantiated track IDs and their corresponding entity, wheres using an `equals` predicate will introduce relations between all objects, which does not seem reasonable in a dynamic domain.

In a previous work we use an extension of the presented model to integrate target information of persons with the tracking data [9]. This work focuses on exploiting a discrete floor plan and examines the sensitivity of the approach to variations of the used plan.

III. Multi-Object Tracking

Standard multi-object tracking algorithms often use object individual single-object trackers like the Kalman filter. The drawback of this multi-object tracking approach is the need of a data association step which assigns the received measurements to the trackers using hard decisions or probabilistic methods [10]. Especially in scenarios characterized by a high object density, the data association is error-prone and degrades the performance of the tracking system, since false associations are irreversible.

A rigorous approach to multi-object tracking is the multi-object Bayes filter proposed by Mahler [4]. The multi-object Bayes filter uses the random finite set statistics to represent the complete environment by a single filter state. In the innovation step of the multi-object Bayes filter, a multi-object likelihood function calculates the affinity between the predicted state set and the received measurement set. Thus, no data association is necessary.

Further, the multi-object Bayes filter allows to integrate state dependent detection probabilities into the filtering algorithm. In Reuter and Dietmayer [5], an approach to calculate state dependent detection probabilities based on the occupancy grid mapping approach [11] is proposed. Thus, it is possible to keep track of an object which is occluded for the sensor for a short period of time. Using constant detection probabilities would lead to a track loss, if an object is not visible to the sensor for a few measurement cycles. We use the state dependent tracking algorithm as a comparison for our final results. But for input into the high-level model, we use state independent object tracking. This produces more track IDs for association, and in particular is less prone to false association, which we cannot correct in the upper stage.

An implementation of the multi-object Bayes filter is possible using Sequential Monte Carlo (SMC) methods [4], [5], [12]. In difference to well known SMC implementations of the standard Bayes filter a particle set, which represents a random finite set using a finite number of state vectors, is used instead of a standard particle. Further, the number of state vectors in the particle set may change at each time step. In case of a SMC implementation, the integration of the mentioned constraints is possible by reducing the weight of a particle set.

Since the multi-object Bayes filter does not perform a measurement to track association, an extraction of the individual objects out of the multi-object posterior density function is necessary, e.g. using the k-means algorithm [13].

IV. Markov Logic Networks

Markov Logic Networks [1] are a member of the family of first-order probabilistic languages [14] and their semantics are based on undirected graphical models (Markov networks). In contrast to propositional models like Bayesian networks and Markov Networks, where every random variable has to be specified explicitly, in first-order models the random variables are relations over objects and the model can be scaled by providing the appropriate number of object constants. Moreover, MLNs allow the specification of dependencies as weighted first-order logical formulas.

Higher weights make those interpretations more likely, in which more groundings of the formula evaluate to true. We will now briefly cover the formal semantics of MLNs.

A *Markov Logic Network* $L = \{(f_1, w_1), \ldots, (f_n, w_n)\}$ for $n \in \mathbb{N}$ is a set of first-order formulas $f_1, \ldots, f_n$ with given weights $w_1, \ldots, w_n \in \mathbb{R}$. Together with a finite set of constants C, they define a probability distribution over all interpretations (or possible worlds). An interpretation maps each grounding of each predicate to a truth value. The interpretation of functions must be fixed. Probabilistic functions can be emulated using predicates. Let $g_C(f)$ be the set of groundings of formula f obtained by replacing the free variables in f by all combinations of constants from C. Given an interpretation x, then $n_{C,i}(x) \stackrel{\text{def}}{=} |\{g \mid g \in g_C(f_i) \text{ and } x \models g\}|$ is the number of groundings of formula f_i that are true under x. Then, the probability distribution $P_{L,C}$ that is defined by the MLN L with constants C is given as

$$P_{L,C}(X = x) \stackrel{\text{def}}{=} \frac{1}{Z} \prod_i \exp\big(w_i n_{C,i}(x)\big), \qquad (1)$$

where i ranges over all formulas in L, and Z is a normalizing constant.

Given a set of constants, a MLN can be converted to a Markov network, where nodes correspond to atoms and each ground formula induces a clique over all nodes whose atoms appear inside this formula. For practical reasons, a sorted (or typed) logical language is used to describe MLNs. Using sorted terms, we can limit the size of the grounded network. Also, in their basic form, MLNs do only allow restricted usage of logical functions. Usually functions are simulated by specially marked predicates, which enforce a functional dependency of one or more arguments on the remaining arguments. We notate functional arguments of predicates by underlining them. Such a predicate can be translated to a multi-valued random variable.

V. Problem Description

We consider an indoor scene which resembles an office setting. The corresponding floor plan is depicted in Figure 1a. A laser range finder is placed in one corner of the main room and provides distance information in a plane about one meter above ground. The beam almost completely covers the main room, but there exists a second room that has virtually no sensor coverage. There is only one entrance to the room complex and the separate room has only a single exit, which is the door to the main room.

There are one to three persons inside the scene simultaneously. Major occlusion caused by static objects, like walls, occurs when people enter the second room. Minor static occlusion can occur near the coffee maker. During the scenes with more than one person, dynamic occlusion occurs when persons are covered by other persons standing between them and the sensor.

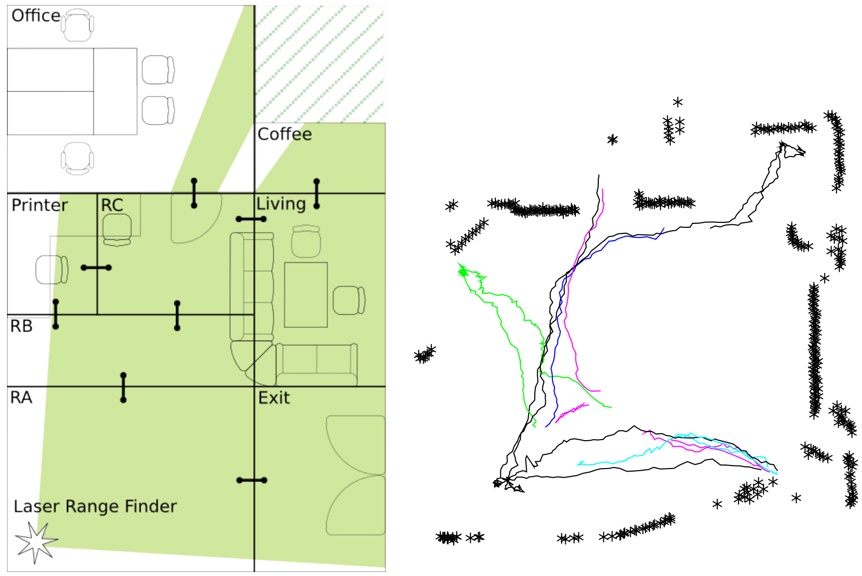
Using only the particle filter-based tracking algorithm to process the output of the laser range finder, problems arise when people produce no measures for an extended period of time because they are inside the separate room or because they are hidden by another person. For shorter occlusion durations it is possible to keep the track of a single occluded person alive for long enough for the person to reappear and re-association is completely handled by the tracking algorithm. If two persons enter the same occlusion area, their estimated positions begin to mix spatially and once they emerge again, re-association becomes more and more arbitrary with increasing occlusion duration. In these scenarios, a direct integration of the Social Force model [15] into the prediction of the persons state of the tracking algorithm may increase the performance of the system [16]. Since the social force model heavily depends on the destinations of the person, a tight integration with a high-level knowledge base, as described in this work, seems promising for such an approach.

Figure 1b shows an example of the tracking results for one of the sequences with three persons. The trajectories are illustrated by solid lines. Since the results are generated without the usage of the state dependent detection probability, the trajectories are interrupted quite often in the area corresponding to region RA, where a dynamic occlusion occurs.

VI. Description of the Model

In this section, we describe the used MLN model and discuss some of the difficulties and design choices we have encountered in its engineering. The complete model is factored into three modules. We begin with a discussion of two concepts that cannot be associated with distinct model parts but influence nearly every aspect – the representation of space and time.

The basic MLN can only represent discrete random variables. There exists an extension of MLNs to continuous variables [17], but no working implementation is available. For this work, we reduce the continuous spatial estimates obtained from the tracking algorithm to a few discrete regions. For ease of modeling and processing of data, we choose a rectangular shape. We do not create a uniform grid, but try to respect functional aspects of the environment concerning the problem. For example, it does not make sense to further split the office into smaller areas if there is no distinction for the sensor (everything is one connected occluded area). In an initial model, we have defined a total of eight regions, which are depicted in Figure 1a and correspond to floor plan D in the evaluation section. For evaluation purposes we have defined a further five region layouts to examine how sensitive the approach is to the choice of discretization. Sticking with a low number of regions also made an exact evaluation of the final model feasible. Depending on the inference approach, there might be no significant overhead



(a) Floor plan (b) Example trajectories

Figure 1: (a) Floor plan of the location used for the experiments. In addition the picture shows the eight discrete regions that are used for the MLN model and how they are connected via the handles. The depicted floor plan omits a static occlusion in the lower right that served as a simulated entrance. The door in the lower right was not opened during the experiments. (b) Example trajectories of a three person sequence: trajectories are illustrated by solid lines in different colors. The black stars are measurements of static objects like walls.

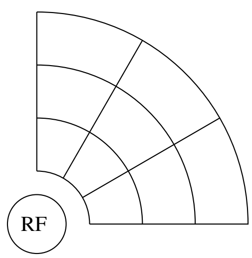


Figure 2: A radial layout of regions that respects functional aspects of the sensor – a laser range finder.

when using a larger state space for the spatial component, e.g. when using sequential Monte Carlo methods. Although, the model engineering may become more intricate when opting for more fine grained regions, the evaluation section shows that the approach is rather insensitive to the concrete layout chosen. Taking the characteristics of the sensor into account, a radial layout as depicted in Figure 2 seems like a promising approach, but this was not investigated.

In order to model dynamic domains, we assign a dedicated time sort, whose constants are elements from the natural numbers. One usually aims to construct a model that fulfills the Markov property, i.e., the state at time $t+1$ only depends on the state at time t. This means that formulas may only contain predicates of at most two different times, which then must be successive. But in the described model, the predicates that represent the association of tracks to persons are not time-indexed, which makes them static. This makes it difficult to apply standard dynamic inference algorithms, which usually assume the Markov property. But a static variable can be considered as a dynamic variable, for which the same value is deterministically enforced in every time step. Fortunately, these static variables are only referenced over a limited period of time — the life-time of a track — so they do not pile up over the course of the complete sequence. For time resolution we have settled for the duration of about one second, which seems like a good compromise between inference complexity and accuracy for the given problem.

The tracking model: The basic functionality for interfacing with the tracking algorithm is provided by a MLN module that contains objects of the sorts `Track` and `Person`. To notate variables of some sort, we use the initial letter of the sort name in lower case. For the sort `Track`, the letter 'm' is used because of the ambiguity with sort `Time`. For both sorts `Track` and `Time`, there exist time-dependent predicates $\mathtt{at}_T : \mathtt{Time} \times \mathtt{Track} \times \underline{\mathtt{Region}}$ and $\mathtt{at}_P : \mathtt{Time} \times \mathtt{Person} \times \underline{\mathtt{Region}}$ that give the current location of a track or person, respectively. The time-independent predicate $\mathtt{a} : \mathtt{Track} \times \underline{\mathtt{Person}}$ associates `Track` objects to `Person` objects. The correspondence of tracks to persons inside the MLN is similar to the correspondence of measurements to tracks inside the tracking algorithm. The output of the tracking algorithm is converted to observations of the $\mathtt{at}_T$

function and an additional completely observed predicate $\texttt{act} : \texttt{Time} \times \texttt{Track}$, whose technical purpose is to indicate the life span of track IDs. The usage of this predicate is sometimes omitted for clarity. The core tracking model then consists of the following three formulas.

$$5 \quad \texttt{at}_T(t, m, \underline{l}) \land \texttt{a}(m, \underline{p}) \Rightarrow \texttt{at}_P(t, p, \underline{l}) \tag{2}$$

$$-3 \quad \texttt{a}(m_1, \underline{p}) \land \texttt{a}(m_2, \underline{p}) \land m_1 \neq m_2 \tag{3}$$

$$2 \quad \texttt{at}_P(t, p, \underline{l}) \land \texttt{at}_P(t + q, p, \underline{l}) \tag{4}$$

Formula 2 probabilistically forces a person to be in the same region as its associated track. By design a track can only be associated to one person at a time because the last parameter of the predicate a is declared functional. Formula 3 probabilistically enforces the association to also be a one-to-one relation. This formula is limited to concurrently instantiated tracks using the act predicate (not listed). And finally Formula 4 puts a bias on people standing still instead of moving every time step.

The floor plan: We use the static predicate $\texttt{adj} : \texttt{Region} \times \texttt{Region}$ to encode the connectedness of the regions. All instances are fully observed and adhere to the given floor plan as indicated in the figures. A single formula forces persons to move between regions only according to the given layout:

$$\infty \quad \texttt{at}_P(t, p, \underline{l_1}) \land \texttt{at}_P(t + 1, p, \underline{l_2}) \Rightarrow \texttt{adj}(l_1, l_2) \tag{5}$$

This formula is deterministic to prevent persons from "teleporting" through the scene, traversing an arbitrary distance in one step. If regions allow for a traversal in less than a second (or one time step), this rule becomes invalid. But since the association of tracks to persons allows for some slack, a person in the model can "catch up" to the location of its real counterpart after some time steps, only violating Formula 2.

The occlusion model: To prevent persons without an associated track from wandering across the scene (since no track influences their current location), we need to express that persons usually have a track unless they are indeed occluded. In our setting both static and dynamic occlusions occur, being caused by walls or other persons, respectively. In the presented approach, only static occlusion information is modeled. This is done by assigning a certain probability to each region that it may contain untracked persons. The probability is larger for areas of high static occlusion, like the separate room. We also assign a higher occlusion to regions that are more likely to be dynamically covered, like the region around the coffee maker. By assigning low occlusion probabilities to central regions that have a good sensor coverage we penalize persons silently slipping past the sensor. Formula 6 is provided once for each region r. The weight w_r is the occlusion value given in the depicted floor plans in Figure 3.

$$w_r \quad \texttt{at}_P(t, p, \underline{r}) \land \neg\exists m : (\texttt{act}(t, m) \land \texttt{a}(m, \underline{p})) \tag{6}$$

There exist two major ways to determine the weights of the probabilistic formulas: Learning from data and elicitation from experts; where for common sense domains, like the one we are dealing with, everyone is usually an expert. Both the learning of weights and the direct specification approaches have been followed in the literature. For the case of our related work, [6] and [8] are employing learning and [7] specify the weights by hand. Due to the limited size and the common sense nature of our dataset we decided to specify the weights ourself.

The approximation to consider the weight as the logarithmic odds of the formula being true [1] can serve as a good starting point, but it only holds as long as formulas do not share predicates. After assigning some reasonable initial values, we iteratively looked at predictions of the model for selected sequences and adjusted the weights if the predictions did not conform with our expectations. For example did we look at the marginal trajectories of persons and compared those with the true trajectories, judging which amount of uncertainty or even wrong prediction is reasonable given the sensor data. This can only be done with automatic parameter training by using some kind of probabilistic ground truth. We worked beginning with the sequences containing fewer persons. It turned out that once we got to the sequences with three persons, the weights did not need much more tweaking.

VII. Preprocessing of Tracking Information

We go on and describe how tracking data is processed for input to the MLN model. After extraction of the individual objects in the multi-object Bayes filter, we obtain a set of single object particles X^t_m for each track ID m and time step t. We then apply two data reduction steps. First, the MLN model works on a coarser time scale of 1.25 steps per second, while the tracking algorithm runs with 12.5 steps per second. We drop the intermediate steps without further processing. A different approach might aggregate them, e.g., by averaging, but this would also distort the meaning of the data, because it cannot be considered a snapshot of the situation anymore.

Depending on the quality of the tracking algorithm and the used object model, there can be many false positive tracks, e.g., when people spread their arms away from their body, crossing the plane of laser beams. To reduce these false tracks, we use the existence probability to eliminate insignificant tracks. It is given by $|X^t_m|/N$; the number of particles for track ID m divided by the total number of particles N. We drop all tracks from a time step whose existence probability is below 0.5. For our test sequences, the output of the tracking algorithm usually contains about thirty tracks per sequence, but only less than ten remain after applying both reduction processes.

For each time step t and each track ID m that survive the described process we add the track as active to our MLN model via observation of the act predicate. We then bin the single object particles into the discrete regions. Most of the time all particles are contained in a single region and we create an observation of the $\mathtt{at}_T$ function. In cases where the particles of a track m spread over several regions we reflect this as probabilistic evidence by adding a formula $(w_l, \mathtt{at}_T(t, m, \underline{l}))$ for each location l and calculating the weight as the logarithmic odds $w_l = \log \frac{p_l}{1-p_l}$, where p_l is the relative frequency of a particle of track m being in region l.

VIII. The Inference Problem

Markov Logic Networks can be seen as template models for undirected graphical models [18]. Their semantics are defined using the ground version of these networks. As such the described MLN represents an undirected version of a dynamic Bayesian network [2]. The effort for exact inference in such models is usually exponential in the number of variables within one time slice, because most variables within one time step become dependent on each other after some steps in most models.

The model described in this work also suffers from this problem. The cause that all variables of a time slice become dependent lies in the probabilistic data association; which is a hard problem at its core. In our case the problem of exact inference is exponential in $m_T + n_P$, where m_T is the maximum number of simultaneous tracks and n_P is the total number of persons in the model. Here m_T stems from the association predicates and n_P are the instantiations of the $\mathtt{at}_P$ predicate for one time step.

For our evaluation we perform exact inference on the model by exploiting context-specific independence [18, pp. 171]. Given an assignment to all association variables, the model factorizes into components for each person and thus becomes tractable. Our largest sequence contained 10 tracks, which results in 3^7 possible associations to three persons after observing the correct association for three initial tracks. After conditioning on the association variables we calculate the partition function for each association using variable elimination along a min-degree variable ordering. This approach is not suited for online filtering. Although there exist generic approaches for filtering in dynamic MLNs [19], [20], for the problem at hand a rao-blackwellized particle filter, which collapses all but the association variables, seems like a good solution [18, pp. 526]. Evaluating the performance of this inference approach on the presented model and comparison to the general approaches is open for future work.

IX. Evaluation

We recorded nine sequences in total; three sequences with one, two and three persons each. The duration of each sequence is about one minute. The course of events is the same among sequences with the same number of persons; the sequences vary during the part where multiple persons walk around the main room in an improvised way.

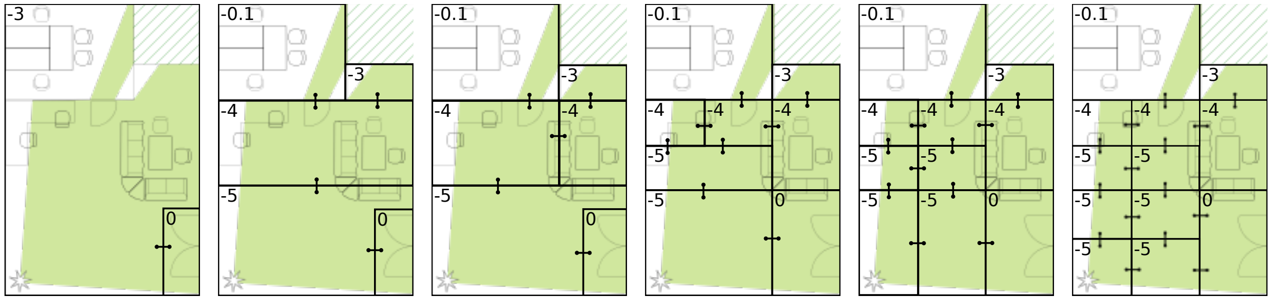
The setups with one person only feature static occlusion caused by the single person staying inside the office for several seconds. This results in its track being reinvented upon entering the main room again. With two persons there is dynamic occlusion, where one person covers the other person. Both persons enter the office together and thus cannot be distinguished once they reappear. Goal information for one person can resolve this issue and we can obtain a good association again. In the scenes with three persons, one person enters the office while the two remaining persons stay inside the main room. Dynamic occlusion occurs while all three persons are walking around the area in front of the sensor. Tracks are lost and recreated often, which can also be observed in Figure 1b inside the area corresponding to region RA. When two persons are simultaneously inside the same occluded area, it is not possible to associate the reappearing tracks to the correct persons just by means of the laser range finder. So we cannot expect perfect scores.

We have created six different floor plan. The fitting of the three core model parameters (Formula 2 - 4) was performed using map D, which was the initial map we used in the experiments. Then five variations of this map, with more and less regions, have been created and the parameters estimated by hand without further fitting. To evaluate the performance of the models, the MLN is instantiated for five persons in every setup, regardless of the number of persons appearing in the scene. For each map and each sequence, we observe the correct association for the first track of each person and evaluate how well we can associate new tracks. In our dataset, the number of tracks that remain unassigned after labeling the starter tracks varies between one and seven.

The results of our evaluation are given in Table I. For each sequence, we compute the most probable association and give the number of false track assignments it performs. In addition, we provide the log-likelihood, which is the natural logarithm of the probability of the true association.

In order to provide a baseline, the number of track confusions and losses of a multi-object Bayes filter with state dependent detection probability ($p_D(x)$) and the ones using a filter with constant detection probability ($p_D = c$) are given. For both filters, the number of persons inside the scene is subtracted from the total number of significant tracks and the result is given in the table. If the tracking algorithm works perfectly, this number will be zero. Remember that the output of the $p_D = c$ filter is used as input for the MLN stage; so this number equates to the number of associations made by the high-level model and thus equals the possible maximum number of false associations.

We compare our results to a multi-object Bayes filter using the set representation, which delivers state of the art performance in domains with high occlusion rates [5]. We observe that the algorithm with the state dependent detec-



(a) Map A (b) Map B (c) Map C (d) Map D* (e) Map E (f) Map F

Figure 3: The region layouts used for evaluation. The numerical values are the occlusion values that are used in Formula 6. The region in the lower right contains the exit and thus has an occlusion value of zero, which means that untracked persons may be inside this region without reducing probability.

Table I: For each of the nine sequences used for evaluation, we give the number of invented tracks minus the number of persons for tracking with constant detection probability ($p_D = c$, maximum number of possible confusions) and with state dependent detection probability ($p_D(x)$), which functions as baseline. For the six floor plans, we give the number of false associations (a_-) of the most probable association and the log-likelihood ($\ln \mathcal{L}$).

		Unassigned Tracks		A		B		C		D		E		F	
Seq	Persons	$\downarrow p_D = c$	$\downarrow p_D(x)$	$\downarrow a_-$	$\uparrow \ln \mathcal{L}$	$\downarrow a_-$	$\uparrow \ln \mathcal{L}$	$\downarrow a_-$	$\uparrow \ln \mathcal{L}$	$\downarrow a_-$	$\uparrow \ln \mathcal{L}$	$\downarrow a_-$	$\uparrow \ln \mathcal{L}$	$\downarrow a_-$	$\uparrow \ln \mathcal{L}$
1532	1	1	1	0	-1.61	0	-0.00	0	-0.00	0	-0.00	0	-0.00	0	-0.00
1640	1	1	1	1	-1.61	0	-0.00	0	-0.00	0	-0.00	0	-0.00	0	-0.00
1737	1	1	1	0	-1.61	0	-0.00	0	-0.00	0	-0.00	0	-0.00	0	-0.00
2056	2	3	2	2	-4.30	2	-2.08	2	-2.08	2	-1.45	0	-0.88	2	-1.58
2207	2	4	4	3	-6.24	1	-3.22	1	-3.22	3	-2.84	3	-2.84	3	-2.85
2329	2	3	3	2	-4.30	2	-2.08	2	-2.08	2	-1.94	2	-1.55	2	-1.73
4628	3	5	1	4	-4.91	0	-1.63	0	-1.61	0	-1.01	0	-1.00	0	-1.00
4734	3	5	3	1	-4.26	2	-4.61	2	-4.62	1	-6.86	2	-7.66	2	-7.42
5306	3	7	4	6	-7.76	4	-8.50	4	-8.50	4	-7.13	4	-7.12	4	-7.11

tion probability reduces the number of unassigned tracks dramatically in the scenarios with three persons, where a lot of short term occlusions occur. In the scenarios with one or two persons, where the long-term occlusions due to static objects dominate, the usage of the state dependent detection probability has nearly no influence on the number of unassigned tracks. We note, that the performance of the MLN model is largely independent of choice of map layout. It is also able to consistently outperform the results achieved by the filter algorithm with state-dependent detection probability on all but the simplest map layouts. Even with map A, the re-association performance of the MLN model is quite good, sometimes surpassing that of more detailed maps. In general the approach is more likely to correctly associate a new track ID to the correct person, if the person's track was lost and recovered inside the same region, since staying still is the most probable behaviour of a person. If a more detailed map places a region boundary between those two points of a person track, the corresponding model receives a penalty.

When interpreting the results one has to think about two further points that are relevant if one was to transfer the model to a real application. First, the inference is done in an exact way, which cannot be sustained for longer sequences containing more track IDs. The used method of inference achieved real-time performance for about 4-5 unassigned tracks (meaning inference took as long as sequence duration). We expect that a properly designed approximate algorithm should be able to deliver comparable performance in an online setting. Secondly, the track-assignment was taken at the end of the sequence, accounting for all evidence. This means that early track associations benefit from a smoothing effect, taking into account later observations. Overall the results do look encouraging and suggest that layering a high-level probabilistic model on top of a tracking filter is a possible way to improve association performance.

X. Conclusion

We have described an approach to solving the data association problem for person tracking with a first-order probabilistic model described using Markov Logic Networks. We showed how to map the output of a regular tracking algorithm into a discrete spatial representation with additional rules describing occlusion properties of the environment.

Especially in scenarios with long-term occlusions, where even the multi-object Bayes filter is not able to continue to track hidden objects, the association using MLN outperforms the solely tracking-based approach when using exact evaluation. On the other hand, a sophisticated tracking algorithm is adequate in scenarios with occlusions of no more than one second and might be able to scale more easily to larger domains.

In future, we plan to integrate additional information, like observations of interactions with static devices or video data, into the first-order model in order to further improve the association performance. It is planned to develop a specialized particle filter algorithm that exploits the context-specific independence present within the described model and is able to scale to larger problem sizes, both in terms of richness of the model and number of concurrent tracks. An application will be the selection of appropriate output devices for human-computer interaction, where it is necessary to be able to estimate the distance of users to computer screens or speakers.

ACKNOWLEDGEMENTS

This work is done within the Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems" funded by the German Research Foundation (DFG).

REFERENCES

[1] M. Richardson and P. Domingos, "Markov logic networks," *Machine Learning*, vol. 62, no. 1-2, pp. 107–136, 2006.

[2] K. Murphy, "Dynamic bayesian networks: Representation, inference and learning," Ph.D. dissertation, University of California, 2002.

[3] V. Gogate and P. Domingos, "Probabilistic theorem proving," in *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence*. AUAI Press, 2011, pp. 256–265.

[4] R. P. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Artech House Inc., Norwood, 2007.

[5] S. Reuter and K. Dietmayer, "Pedestrian tracking using random finite sets," in *Proceedings of the 14th International Conference on Information Fusion*, 2011, pp. 1–8.

[6] A. Sadilek and H. Kautz, "Recognizing multi-agent activities from GPS data," in *Proceedings of 24th AAAI Conference on Artificial Intelligence*, 2010, pp. 1134–1139.

[7] S. Tran and L. Davis, "Event modeling and recognition using markov logic networks," *Computer Vision–ECCV 2008*, pp. 610–623, 2008.

[8] P. Singla and P. Domingos, "Entity resolution with markov logic," in *Data Mining, 2006. ICDM'06. Sixth International Conference on*, 2006, pp. 572–582.

[9] T. Geier, S. Reuter, K. Dietmayer, and S. Biundo, "Goal-based person tracking using a first-order probabilistic model," in *Proceedings of the Nineth UAI Bayesian Modeling Applications Workshop*, 2012.

[10] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House Publishers, 1999.

[11] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[12] H. Sidenbladh and S.-L. Wirkander, "Tracking random sets of vehicles in terrain," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2003, p. 98.

[13] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 2006.

[14] R. de Salvo Braz, E. Amir, and D. Roth, "A survey of first-order probabilistic models," in *Innovations in Bayesian Networks*, ser. Studies in Computational Intelligence, D. Holmes and L. Jain, Eds. Springer, 2008.

[15] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *PHYSICAL REVIEW E*, vol. 51, pp. 4282–4286, 1995.

[16] S. Reuter and K. Dietmayer, "Adapting the state uncertainties of tracks to environmental constraints," in *Proceedings of the 13th International Conference on Information Fusion*, 2010, pp. 1–7.

[17] J. Wang and P. Domingos, "Hybrid markov logic networks," in *Proceedings of the 22th AAAI Conference on Artificial Intelligence*, 2008, pp. 1106–1111.

[18] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[19] T. Geier and S. Biundo, "Approximate online inference for dynamic markov logic networks," in *International IEEE Conference on Tools with Artificial Intelligence*, 2011, pp. 764 –768.

[20] T. Papai, H. Kautz, and D. Stefankovic, "Slice normalized dynamic markov logic networks," in *Proceedings of the Conference on Neural Information Processing Systems*, 2012.