Better Models For People Tracking

Matthias Luber Gian Diego Tipaldi Kai O. Arras

***Abstract*— People tracking is a key component for robots operating in populated environments. Previous works have employed different filtering and data association techniques for this purpose that typically rely on a set of generic assumptions on target behavior and detector characteristics. In this paper, we focus on these assumptions rather than the tracking approach itself and show that with informed models, people tracking can be made substantially more accurate without compromising efficiency. Concretely, we present better, human-specific models for the occurrence of new tracks, false alarms, track occlusions, and track deletions. In the experiments with a large-scale outdoor data set collected with a laser range finder, the models and combinations thereof are experimentally compared using a multi-hypothesis baseline tracker and the CLEAR MOT metrics. The results show how some models selectively improve tracking performance at the expense of other measures. The final combination is then able to resolve the trade-offs, leading to a reduction of data association errors by more than a factor of two at the same cost.**

I. Introduction

As robots enter domains in which they interact and cooperate closely with humans, people tracking is becoming a key technology for many research and application areas in robotics and related fields.

The task has been addressed with a variety of general-purpose target tracking techniques that employ different filtering and data association schemes. Typically, these systems make generic assumptions about the target behavior and the detector characteristics. But for people as targets, some of these assumptions are overly simplistic and ignore important information that is available. For example, new tracks are often assumed to be uniformly distributed over the sensor field of view. But the way how people move is often given by static environmental constraints that can be learned. Indoors, for instance, doors or convex corners are typical places where people appear. The same place-dependency applies for the behavior of a detector. Regions of clutter and complex background produce false alarms more likely than in open space, making a spatially uniform model a poor approximation.

The techniques that have been employed for people tracking in past include Kalman filters (KF) and nearest-neighbor data association [1], [2], particle filters and a sample-based variant of the Joint Probabilistic Data Association filter (JPDAF) [3], the KF-based multi-hypothesis tracking filter (MHT) [4], [5], [6], or a KF-based tracker in which the data association is formulated as a Minimum Description Length problem and solved using Quadratic Boolean Programming [7]. Tracking people becomes particularly challenging if the targets are identical in appearance which is typically the case for tracking using radar or laser range finders. With a good, target-specific appearance model, many hard tracking problems such as dealing with occlusions and interactions of tracks, becomes much easier to cope with. For this reason, visual tracking systems, where rich appearance information is available, can achieve good results with nearest-neighbor data association filters as in [8] using a set of independent particle filters. However, in this paper, we assume targets to have identical appearance.

All authors are with the Social Robotics Lab, Department of Computer Science, University of Freiburg, Germany {luber,tipaldi,arras}@informatik.uni-freiburg.de.

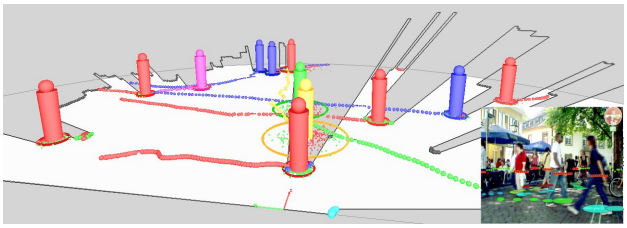

Fig. 1. Twelve of 162 tracks from the outdoor experiment. The cylinders show the estimated positions of the pedestrians, the colored dots illustrate their past trajectories.

The paper is structured as follows. The next section reviews related work. Section III introduces the theory of the proposed models. Section IV provides a short overview of the multi-hypothesis tracking approach and describes how the models are integrated into this framework. Section V presents the experimental results followed by the conclusions in section VI.

II. Related Work

Every tracking system needs to deal with new tracks, false alarms, missed detections, occlusions and track terminations. We review the people tracking literature with respect to how these events have been modeled.

Schulz *et al.* [3] propose a local occlusion grid to determine the probability of a track being occluded or a measurement being missed in a sample-based JPDAF framework. Taylor *et al.* [4] track legs of a single person in laser data using a MHT. Based on the relative positions of legs to each other, the occlusion probability is approximated with a piecewise linear model. Arras *et al.* [6] reformulate the MHT expressions to make the occlusion probability an explicit parameter. Then they track multiple people by separately tracking legs and adapt the occlusion probabilities of tracks as soon as tracks are recognized to belong to a person. In Katz *et al.* [9] probabilistic occlusion checking is used to improve

the robustness of a motion detection algorithm. The occlusion probabilities are computed by a sample-based visibility check for each track. A similar model has also been used by Mucientes *et al.* [5]. For the purpose of vision-based multi-person tracking Ess *et al.* [10] model occlusions in a occlusion grid map, keeping tracks alive that are known to be hidden by other tracks and static objects as a hard decision.

With the exception of [6], all these works compute the final occlusion probability on a per-track basis by a geometric visibility test that determines if or how far a track is 'in the shadow' of other tracks or static objects. In [4] this is realized using a piecewise linear model, all others use samples to this end.

For track terminations or deletions, one can assume a constant deletion probability as in the regular MHT approach [11]. Counting the number of consecutive non-confirmations of a track and deleting it when it exceed a threshold has been done e.g. in [8]. This simulates the decrease in probability of detecting a target that could not be assigned to an observation in several consecutive frames. Mucientes *et al.* [5] track clusters of people in laser range data, modeling the probability of deleting a track from a cluster with an exponential decay function. In [3], the same principle was realized using a discounted average weight of the particles that automatically decreases when tracks are no longer confirmed. Weak tracks in this sense are then deleted if a mismatch with the number of observations is encountered. In Lin *et al.* [12] the track score based on a likelihood ratio of deleting or not deleting the track is computed. If this score falls below a given threshold the track is deleted.

The arrival of new tracks is modeled by Schulz *et al.* [3] as a Poisson process with constant rate over time and a uniform distribution over space. The same assumptions are taken in the regular MHT approach. A simple form of place-dependency has been realized by Breitenstein *et al.* [8], a visual surveillance scenario with a static camera, where a predefined area around the border of the image has been manually put to describe the region where new tracks may appear. It is assumed that no new tracks arrive in the center of the image.

The false alarms model that is employed in most related works, e.g. in Khan *et al.* [13], predicts spurious measurements uniformly over the sensor field of view. This is also the assumption in the original MHT.

In our previous work we already addressed two of these models [14]. We learned place-dependent Poisson rate functions that describe the occurrence of false alarms and new tracks. The approach overcomes the assumption that these events are uniformly distributed in space and is able to model that people typically appear at specific locations in the environment and that false alarms occur more likely in places with clutter.

In this paper we combine this approach with a sample-based occlusion model that incorporates geometric information from the scene and a deletion model that assumes exponentially distributed interarrival times of observations. We extend the state of the art where these models have been considered only in isolation by a systematic experimental review of the effects of each model and their combinations. We carry out large-scale experiments with a challenging data set, collected in the city center of Freiburg and compare the different models using the recently proposed CLEAR MOT metric [15]. The insights gained in this paper are valid for people tracking in general regardless the sensor modality, the filtering approach, or the space in which targets are represented.

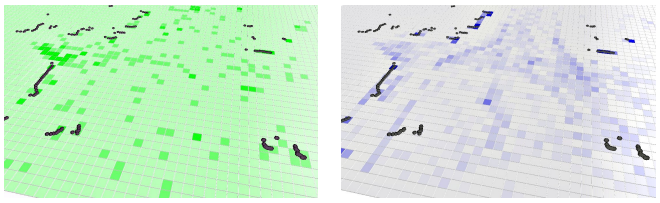

Fig. 2. Learned spatial priors in the environment of the data set used in the experiments. The probability distribution of new track events is shown on the left, the distribution of false alarm events is shown on the right. Local maxima of the new track distribution denote locations where people appear often in the sensor field of view. High probability regions in the false alarm distribution denote clutter and areas in which detector failures are more likely.

III. Models for People Tracking

This section introduces the different models that we consider in our comparison.

A. New Track and False Alarm Models

During tracking, there are situations where a sensor observation cannot be explained by any of the current tracks. The measurement is therefore either spurious (a false alarm or false positive) or a new target that entered the sensor field of view. It is typically impossible to determine which of the two interpretations is correct from a single scan or image. Instead, probabilities for both events can be computed, and if the tracking framework is able to integrate data association and interpretations over time, decisions can be taken in a delayed fashion. In order to compute probabilities, models are required that predict how often and where new tracks and false alarms occur. As mentioned in the previous section, the general assumption is that new tracks and false alarms occur both uniformly over the sensor field of view at rates that follow a Poisson distribution.

This assumption may be valid for traditional setups in target tracking in which airborne targets are sensed by an upwards looking radar or setups that do not use a target detector. For people, however, this model does not account for the place-dependent character of human behavior and the place-dependent character of visual or range-based people detectors. People typically appear, disappear, walk and stand at specific locations that correspond, for instance, to doors, elevators, or convex corners. A similar insight is true for people detectors that

are more prone to false positives in areas of background clutter and at locations of objects with a target-like appearance, leading to systematic misdetections.

We address these issues by learning a spatial Poisson process that predicts both new track and false alarms using a spatio-temporal rate function. As the model itself, the theory and the way learning is done in this case of a Poisson process has been recently presented in [14], we summarize the main ideas. The model is a non-homogeneous spatial Poisson process with rate functions $\lambda_{new}(\mathbf{x}, t)$ and $\lambda_{fal}(\mathbf{x}, t)$ that are functions of space $\mathbf{x} \in \mathbb{R}^2$ and time t. These functions are learned in a grid approximation (making it a piecewise homogenous process). For each grid cell, the rates can be learned from tracking targets and counting each type of events (new tracks or false alarms) that occurs in that cell. We use Bayesian parameter learning to obtain a rate estimate that is better suited for our task than the corresponding maximum likelihood estimate. As it has been shown in [14], estimating a Poisson rate becomes a parameter estimation problem of a Gamma distribution, which, after some simplifications, leads to a straightforward expression to count in a grid. Concretely, the rate estimate $\widehat{\lambda}_{\text{Bayes}}$ for a given type of events is obtained by

$$\widehat{\lambda}_{\text{Bayes}} = \frac{\#\text{positive events} + 1}{\#\text{observations} + 1} \,. \tag{1}$$

We learn events using a baseline tracker with learned fixed parameters that each time when the best hypothesis contains a new track at a given position, the corresponding cell is updated accordingly.

Given a learned rate function $\lambda_{new}(\mathbf{x}, t)$ for new track events (for false alarms the same procedure applies), we obtain a probability distribution by normalization

$$p_{new}(\mathbf{x}) = \frac{\lambda_{new}(\mathbf{x}, t)}{\lambda_{new}(t)}, \tag{2}$$

with $\lambda_{new}(t) = \int_V \lambda_{new}(v, t)\, dv$ where V is the sensor field of view. Figure 2 shows the learned rate functions for both types of events for the environment in which we collected the data set.

B. Occlusion Model

When an existing track cannot be confirmed by an observation, the system has to decide if the track is still in the surveilled area or not. In such a situation we can distinguish four cases: occlusion, interaction, missed detection, and track termination. This subsection deals with a model for the former three cases, track termination is considered in the next subsection.

Occlusions occur when (far apart) targets or static objects occult other targets. Interactions are situations in which close targets interact with each other, potentially changing their behavior, and appear as a single observation. These events are different from detection failures that typically happen with a probability that does not depend on the past, while occlusions and interaction usually occur in an interval of time. In fact, while missed detection can be handled well by data association techniques with delayed decision taking such as MHT [16] or Markov chain methods [17], lengthy interactions and occlusions are notoriously challenging when targets are identical in appearance.

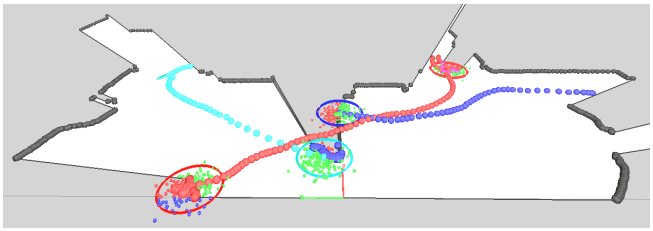

Fig. 3. Visualization of the occlusion model for an example scene. Black dots mark the laser end points connected by a black line that indicates the border of the visible area. State uncertainties and trajectories of four persons are shown with colored circles and dots. The small dots are particles drawn from the state predictions. They are colored in green when they are inside the visible area and red when they are occluded. Blue particles fall outside the sensor field of view which is limited to 180 degrees. The occlusion probabilities of the tracks are, from left to right, 0.42, 0.16, 0.69, and 0.47.

The model proposed in this section aims to explain occlusions by the geometry of the scene, i.e. when people are hidden by other people or objects. Two aspects need to be considered. First, how to encode the visibility of the scene in some representation M_{occ}. Authors have approached this with either simple visibility checks [4], [9] or more complex occlusion maps [3], [10]. In our case, M_{occ} is given by the contour derived from the laser points of the current scan.

The second aspect is the knowledge about the target position. Unlike [10] where only the first moments in a non-probabilistic manner are considered, the occlusion probability should also depend on the uncertainty of the expected target position. Thus, the targets are predicted given their past location $\mathbf{x}_{t-1}$ according to the motion model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$. Following a sample-based approach similar to [5], [9], we determine the occlusion probability for a track $\mathbf{t}_i$ as

$$p_{occ}(\mathbf{t}_i \mid M_{occ}) \approx \frac{1}{N} \sum_{j=1}^{N} p_{occ}(x^{(j)} \mid M_{occ}), \tag{3}$$

where N is the number of samples and $x^{(j)}$ is drawn from $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ of track $\mathbf{t}_i$. Figure 3 shows an example scene and the behavior of the occlusion model.

C. Deletion Model

When targets disappear from the sensor field of view, their tracks need to be declared as obsolete. Otherwise they inflate the system and unnecessarily increase the level of data association ambiguity. As discussed in section II, the common approaches are either a constant deletion probability as in [11] or to update some score for tracks that have not been confirmed through a sequence of steps and delete them if a threshold is exceeded [5], [8], [3], [12]. While both approaches have been shown

to be practical in the past, these models consider track deletions in isolation and not jointly with track occlusions although they both try to explain non-detections of existing tracks. Therefore, the only alternative reasons for tracks being not confirmed in such an isolated model are missed detections. Similar to the approach in [5], we model the deletion probability of a track with an exponential function to simulate the decay in the probability of detecting it when it has not been matched for several consecutive iterations. More formally, let $t - t_0$ be the number of consecutive timesteps that target $\mathbf{t}_i$ has not been observed, the deletion probability is defined as

$$p_{del}(\mathbf{t}_i) = 1 - exp\left(-\frac{t - t_0}{\lambda_{del}}\right), \tag{4}$$

where λ_{del} is the speed of the decay process. The theoretical insight of this model is that Eq. 4 represents the cumulative density function of an exponential distribution with parameter $1/\lambda_{del}$. This exponential distribution thus represents the probability distribution of the interarrival times of observations – following a Poisson process model for observations. The deletion probability is then the natural result for the probability of not having observed the track after a certain duration.

IV. Multi-Hypothesis Tracking of People

In this section, we show how the presented models are integrated into the tracking framework.

The MHT algorithm hypothesizes about the state of the world by considering all statistically feasible assignments between measurements and tracks and all possible interpretations of measurements as false alarms or new track and tracks as matched, occluded or obsolete. Thereby, the MHT handles the entire life-cycle of tracks from creation and confirmation to occlusion and deletion. In the original paper by Reid [11], measurements can be interpreted as matches with existing tracks, new tracks, or false alarms. Tracks are interpreted as *detected* when they match with a measurement or *not detected*. Cox *et al.* [16] extend this framework with the interpretation of tracks as *deleted* and Arras *et al.* [6] extend the MHT framework to multiple track interpretations including *occlusions*.

Formally, let Ω_l^t be the l-th hypothesis at time t and $\Omega_{p(i)}^{t-1}$ the parent hypothesis from which Ω_l^t was derived. Let $Z(t) = \{\mathbf{z}_i(t)\}_{i=1}^{m_t}$ be the set of m_t measurements which in our case is the set of detected people in the laser data. Let further $\psi_i(t)$ denote a set of assignments which associates predicted tracks to measurements in $Z(t)$ and let Z^t be the set of all measurements up to time t. Starting from a hypothesis of the previous time step, called a parent hypothesis $\Omega_{p(i)}^{t-1}$, and a new set $Z(t)$, there are many possible assignment sets $\psi(t)$, each giving birth to a child hypothesis that branches off the parent.

The probability of the hypothesis is recursively calculated as the product of a normalizer η, a measurement likelihood, an assignment set probability and the parent hypothesis probability

$$\begin{aligned} p(\Omega_l^t \mid Z^t) &= \eta \cdot p(Z(t) \mid \psi_i(t), \Omega_{p(i)}^{t-1}) \\ &\quad p(\psi_i(t) \mid \Omega_{p(i)}^{t-1}, Z^{t-1}) \cdot p(\Omega_{p(i)}^{t-1} \mid Z^{t-1}), \end{aligned} \tag{5}$$

where the last term is the probability of the parent hypothesis which is known from the previous iteration.

For the measurement likelihood, it is assumed that a measurement $\mathbf{z}_i(t)$ associated to track $\mathbf{t}_j$ has a Gaussian pdf centered on the measurement prediction $\hat{\mathbf{z}}_j(t)$ with innovation covariance matrix $S_{ij}(t)$, $\mathcal{N}(\mathbf{z}_i(t)) := \mathcal{N}(\mathbf{z}_i(t)\,; \hat{\mathbf{z}}_j(t), S_{ij}(t))$. The regular MHT now assumes that the pdf of a measurement belonging to a new track or false alarm is uniform in V, the sensor field of view, with probability V^{-1}. Thus

$$p(Z(t) \mid \psi_i(t), \Omega_{p(i)}^{t-1}) = V^{-(N_{fal}+N_{new})} \cdot \prod_{i=1}^{M_t} \mathcal{N}(\mathbf{z}_i(t))^{\delta_i}, \tag{6}$$

with N_{fal} and N_{new} being the number of measurements labeled as false alarms and new tracks, respectively. δ_i is an indicator variable being 1 if measurement i has been associated to a track, and 0 otherwise.

The original expression for the assignment set probability can be shown to be [6]

$$\begin{aligned} p(\psi_i(t) \mid \Omega_{p(i)}^{t-1}, Z^{t-1}) &= \eta' \cdot p_{det}^{N_{det}} \cdot p_{occ}^{N_{occ}} \cdot p_{del}^{N_{del}} \\ &\quad \lambda_{new}^{N_{new}} \cdot \lambda_{fal}^{N_{fal}} \cdot V^{(N_{fal}+N_{new})}, \end{aligned} \tag{7}$$

where N_{det}, N_{occ}, and N_{del} are the number of matched, occluded and deleted tracks, respectively. The parameters p_{det}, p_{occ}, and p_{del} denote the constant probabilities of matching, occlusion and deletion subject to $p_{det} + p_{occ} + p_{del} = 1$. Again, the regular MHT assumes that the number of new tracks N_{new} and false alarms N_{fal} both follow a fixed rate Poisson distribution with expected number of occurrences $\lambda_{new}V$ and $\lambda_{fal}V$ in the observation volume V.

A. Integration into the MHT

The integration of the presented models is particularly simple in the case of the MHT. The fixed rates $\lambda_{new}V$ and $\lambda_{fal}V$ are substituted by the learned and normalized rate functions $\lambda_{new}(\mathbf{z}_i(t), t)$ and $\lambda_{fal}(\mathbf{z}_i(t), t)$ from Eq. 2 where $\mathbf{z}_i(t)$ is the position of observation i at time t. The track dependent terms p_{occ} and p_{del} also go directly into the final expression for a hypothesis probability

$$\begin{aligned} p(\Omega_l^t \mid Z^t) \;=\; & \eta'' \cdot p(\Omega_{p(i)}^{t-1} \mid Z^{t-1}) \cdot p_{det}^{N_{det}} \\ & \prod_{i=1}^{N_t} \left[(\mathcal{N}(\mathbf{z}_i(t))^{\mu_i} \cdot (p_{occ}^i)^{\omega_i} \cdot (p_{del}^i)^{\delta_i} \right] \\ & \prod_{i=1}^{M_t} \left[\lambda_{new}(\mathbf{z}_i(t), t)^{\nu_i} \cdot \lambda_{fal}(\mathbf{z}_i(t), t)^{\phi_i} \right], \end{aligned} \tag{8}$$

where μ_i, ω_i, and δ_i are indicator variables whether a track is matched to a measurement, occluded or marked

as deleted, respectively. The indicator variables ν_i and ϕ_i select when a measurement is declared as a new track or false alarm. The probabilities p^i_{occ} and p^i_{del} are calculated with the occlusion and deletion models described above. For all tracks p_{det}, p^i_{occ} and p^i_{del} have to be normalized to sum up to one.

V. Experiments

The experiments were carried out on a large, unscripted outdoor data set collected in the city center of Freiburg during a regular work day. It consists of 55,475 frames recorded over 25 minutes. The sensor used for collecting the data is a fixed laser range finder with an angular resolution of 0.5 degree, mounted at a height of 0.85 meter. We chose a fairly busy crossing of alleys that is used by individuals, couples, groups of people, bicycles, cars, people in wheelchairs, subjects on skates and person-shaped static obstacles that all undergo countless occlusions (see also Fig. 1 and Fig. 4). We manually labeled 10,000 frames with 162 person tracks to determine the ground truth detections and ground truth data associations. The data sets are available on the author's home page.

A fixed parameter MHT serves as baseline in our experiments. The parameters for detections, occlusions, deletions and the fixed rates for false alarms and new tracks have been learned from another training data set with 95 tracks over 28,242 frames. In detail, $p_{det} = 0.7$, $p_{occ} = 0.27$, $p_{del} = 0.03$, $\lambda_{new} = 0.0002$, and $\lambda_{fal} = 0.005$, respectively. As person detector we use the boosted feature approach presented in [18] which computes a set of geometrical and statistical features for groups of laser points and creates a strong classifier based on decisions stumps as weak learners. This classifier has also been learned from a separate training set.

For the spatial Poisson process model for new tracks and false alarms, we use a 30 *cm* cell resolution. The rate functions are learned using the baseline tracker and the labeled detection ground truth. The occlusion model uses 200 samples per track, the deletion model a decay parameters of $\lambda_{del} = 20$. All experiments are conducted with $N_{hyp} = 300$ hypotheses. N_{hyp} has been varied between 50 and 1050 to verify the behavior in all runs which was found to be stable.

To compare the impact of the presented models onto the tracking performance we first test the individual models against the baseline tracker and then the combinations that makes sense. The accuracy of the resulting strategies is then measured using the CLEAR MOT metrics proposed by Bernardin *et al.* [15]. The metric counts three numbers with respect to the ground truth that are incremented at each frame: misses (missing tracks that should exist at a ground truth position, FN), false positives (tracks that should not exist, FP), and mismatches (track identifier switches, ID). The latter value quantifies the ability to deal with occlusion events that typically occur when tracking people. From these numbers, two values are determined: MOTP (avg. metric distance between estimated targets and ground truth) and MOTA (avg. number of times of a correct tracking output with respect to the ground truth). We ignore MOTP as it is based on a metric ground truth of target positions which is unavailable in our data. Note that, for people tracking, the three error types, FN, FP, and ID, are not equally important. The key challenge of a people tracker, according to our experience, is to maintain the identity of tracks through occlusions, misdetections, interactions and maneuvers. Delayed track termination of people that leave the field of view or delayed track creation are, compared to this, less relevant aspects.

Model(s)	FN	FP	ID	MOTA	*Hz*
baseline	2979	3996	290	0.729	12.6
occlusion	2049	6018	223	0.692	11.7
deletion	1957	6843	225	0.665	11.6
new track	3422	3112	228	0.735	14.8
false alarms	3426	3271	270	0.742	15.1
occ + del	**1821**	6903	211	0.668	10.4
new + false	3761	**1971**	210	0.779	15.4
all	2276	2563	**133**	**0.817**	12.4

TABLE I
CLEAR MOT results on the Freiburg city center data set.

The results of the comparison are given in Table I which contains the CLEAR MOT values and the average cycle time in *Hz* for the baseline tracker, the isolated models and their combination.

A. Discussion of the results

The occlusion and deletion models explain tracks not assigned to any observation due to missing detections from detector failures or occlusions. The results show that the models are able to fill such detection gaps and reduce the number of misses (FN) to 1821 from 2979 of the baseline that wrongly terminates many of these tracks. They also reduce the number of mismatches (ID) to 211 where the baseline incorrectly recreates new tracks. However, this comes at the expense of a higher number of false positives (6903) as incorrect detections (e.g. trees) are also retained more persistently, increasing the FP count at every frame.

The new track and false alarm models explain observations not assigned to any track. The learned place-dependent rate functions for the appearance of new tracks and false alarms enable them to bring down the values for false positives (FP) since the functions implement a form of background learning that removes systematic misdetections (1971 vs. 3996 for the baseline). They also improve the number of mismatches (to 210) as during data association, the system can take better, place-dependent decisions on track creations e.g. after occlusion events. This comes at the expense of a conservative track creation behavior for sporadically recognized targets that enter the field of view (FN increases to 3761).

In summary, the isolated models are only able to make selective improvements, trading off the different

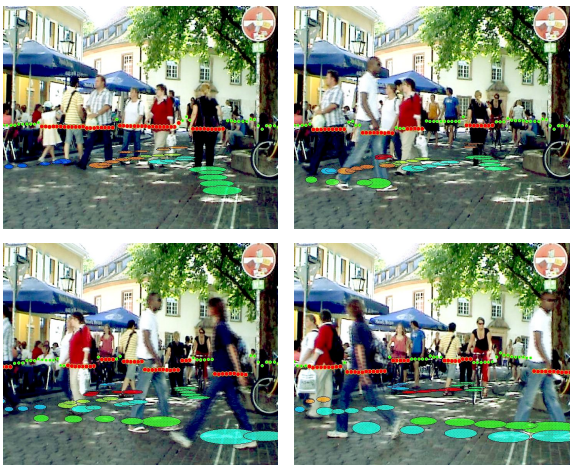

Fig. 4. Four images of the experiment carried out in the city center of Freiburg. The information of the laser range finder and the tracking system were projected onto a recorded video sequence. The images show the tracking results at $t = 335, 353, 365$ and 381. Laser points are shown as green dots (background) or red dots (detected people). Colored ellipses show the traces of the people tracks.

performance aspects. The combination of all models is able to resolve these trade-offs to most parts. While the FN and FP numbers are higher compared to the combination of the specialized models, the most relevant figure, ID, is reduced to 133 which is an improvement of more than a factor of two over the baseline.

This encouraging result comes at practically no additional costs. Compared to the cost of the MHT data association machinery, the computational effort for all models, including the occlusion model that employs sampling, are negligible. This is particularly true for the new track and false alarm model that replace a fix Poisson rate by a learned function, simply realized by a lookup into a grid. The cycle time differences in Table I are due to the behavior differences of the tracking system caused by the models. For instance, more false positive tracks inflate the system and raise the level of data association ambiguity, which in turn, leads to a slower tracker.

VI. Conclusions

In this paper we presented and compared informed target and detector models for the task of tracking people. The models overcome the rather generic assumptions in related work and have been shown to significantly improve tracking performance.

In the experiments using a large-scale outdoor data set and a recently introduced performance metric, we systematically evaluated the impact of the models individually and in combination. We found that the combined application of all models performs best as it is able to resolve the trade-offs introduced by some of the models applied in isolation. The combination leads to an improvement in terms of track identity confusions – the aspect that is most relevant for people tracking – of more than a factor of two at no additional cost. This has been achieved by integrating a set of rather easy-to-use models leaving the much more complex filtering, data association or target detector machineries unaltered.

Acknowledgment

This work has partly been supported by the German Research Foundation (DFG) under contract number SFB/TR-8.

References

[1] B. Kluge, C. Köhler, and E. Prassler, "Fast and robust tracking of multiple moving objects with a laser range finder," in *Proc. of the Int. Conf. on Robotics & Automation (ICRA)*, Seoul, Korea, 2001.

[2] A. Fod, A. Howard, and M. Mataríc, "Laser-based people tracking," in *Proc. of the Int. Conf. on Robotics & Automation (ICRA)*, Washinton, DC, USA, 2002.

[3] D. Schulz, W. Burgard, D. Fox, and A. Cremers, "People tracking with a mobile robot using sample-based joint probabilistic data association filters," *International Journal of Robotics Research (IJRR)*, vol. 22, no. 2, pp. 99–116, 2003.

[4] G. Taylor and L. Kleeman, "A multiple hypothesis walking person tracker with switched dynamic model," in *Proc. of the Australasian Conf. on Robotics & Automation*, Canberra, Australia, 2004.

[5] M. Mucientes and W. Burgard, "Multiple hypothesis tracking of clusters of people," in *Proc. of the Int. Conf. on Intelligent Robots and Systems (IROS)*, Beijing, China, 2006.

[6] K. O. Arras, S. Grzonka, M. Luber, and W. Burgard, "Efficient people tracking in laser range data using a multi-hypothesis leg-tracker with adaptive occlusion probabilities," in *Proc. of the Int. Conf. on Robotics & Automation (ICRA)*, Pasadena, California, USA, 2008.

[7] B. Leibe, K. Schindler, N. Cornelis, and L. V. Gool, "Coupled object detection and tracking from static cameras and moving vehicles," *IEEE Trans. on Pattern Analysis & Machine Intell.*, vol. 30, no. 10, 2008.

[8] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. V. Gool, "Robust tracking-by-detection using a detector confidence particle filter," in *IEEE Int. Conf. on Comp. Vis. (ICCV)*, Kyoto, Japan, 2009.

[9] R. Katz, J. Nieto, and E. M. Nebot, "Probabilistic scheme for laser based motion detection," in *Proc. of the Int. Conf. on Intelligent Robots and Systems (IROS)*, Nice, France, 2008.

[10] A. Ess, K. Schindler, B. Leibe, and L. van Gool, "Improved multi-person tracking with active occlusion handling," in *Proceedings of the IEEE ICRA 2009 Workshop on People Detection and Tracking*, Kobe, Japan, 2009.

[11] D. B. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, 1979.

[12] L. Lin, Y. Bar-Shalom, and T. Kirubarajan, "New assignment-based data association for tracking move-stop-move targets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 2, pp. 714–725, Apr 2004.

[13] Z. Khan, T. Balch, and F. Dellaert, "MCMC data association and sparse factorization updating for real time multitarget tracking with merged and multiple measurements," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, December 2006.

[14] M. Luber, G. D. Tipaldi, and K. O. Arras, "Place-dependent people tracking," in *Proc. of the Int. Symposium of Robotics Research (ISRR)*, Lucerne, Switzerland, 2009.

[15] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: the clear mot metrics," *J. Image Video Process.*, vol. 2008, pp. 1–10, 2008.

[16] I. J. Cox and S. L. Hingorani, "An efficient implementation of reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)*, vol. 18, no. 2, pp. 138–150, 1996.

[17] S. Oh, S. Russell, and S. Sastry, "Markov chain monte carlo data association for general multiple-target tracking problems," in *Proc. 43rd IEEE Conf. Decision and Control*, Atlantis, Bahamas, 2004.

[18] K. O. Arras, O. Martínez Mozos, and W. Burgard, "Using boosted features for the detection of people in 2d range data," in *Proc. of the Int. Conf. on Robotics & Automation*, Rome, Italy, 2007.