Department of Mechanical Engineering
Control Systems Technology Group

Multiple Hypothesis People Tracking in Occluded Environments

Bachelor Thesis

Thijs Wester
1237030

Supervisor:
dr. ir. René van de Molengraft

Coordinators:
Josja Geijsberts
Lars Janssen

Eindhoven, June 2020

Abstract

The field of applications for domestic robots is increasing at a rapid pace. However, their operation in crowded environments is challenging and requires information about humans' positions and poses for reasons such as motion planning, collision avoidance, and interaction. For this purpose, several multiple target tracking techniques were developed. The multiple hypothesis tracker is preferred, but it is computationally costly. Furthermore, object tracking in the presence of occlusion events is still challenging. In this thesis, an extension to the probabilistic multiple hypothesis anchoring algorithm is proposed to improve the handling of occlusions. Occluded areas are identified based on a static map and the object's behavior model is changed accordingly. In simulations, it is shown that the extension significantly increases the representation of the world for scenarios with occlusion events. This could be used in future work to reduce the computational cost of the multiple hypothesis tracker.

Contents

Chapter 1

Introduction

Domestic robots are robots that operate in dynamic and cluttered domestic environments. Whereas the first applications were mainly intended for entertainment, their functionality is increasing at a rapid pace. To boost the research and development of domestic robots, the RoboCup@Home competition[1] was launched in which teams from all over the world compete against each other in so-called challenges.

Since domestic robots operate in environments where humans are present, information about humans' positions and poses is required for reasons such as motion planning, collision avoidance, social awareness, and interaction [1, 2]. Therefore, an important challenge in the RoboCup@Home competition is the detection and tracking of people. Although most tracking techniques already originate from the 20th century for air- and water-borne objects in radar and sonar data, people tracking is a relatively new domain of research containing its specific challenges such as high occlusion probability and non-linear human motion behavior [3]. Furthermore, to achieve full autonomy in a robot, it should be independent of external sensors and computers. Therefore, the tracking system needs to be computational efficient [4].

Tracking techniques vary in optimality, complexity, and robustness [5]. For deterministic tracking, a Multiple Hypothesis Tracker (MHT) approach is assumed to be optimal, but at a high computational cost [6, 7]. In [6, 8, 9] it was found that in dynamic environments sub-optimal but computational efficient techniques such as the Nearest Neighbor (NN) approach perform equally well compared to an MHT approach in real-time applications. Due to the time constraint, the generation, maintenance, and evaluation of hypotheses are impeded to an extent at which the MHT approach does not have added value anymore. This illustrates the need for effective management of the growth of the hypothesis tree.

Most current techniques that limit this growth focus on pruning. However, it would be more efficient to limit the generation of hypotheses instead of pruning generated hypotheses. In [10] it is shown that improving the assumptions on which the MHT approach relies results in a better representation of the world due to which fewer hypotheses are needed. The authors replaced the assumption that new tracks and false alarms are uniformly distributed by place-dependent object behavior. This led to improved tracking results at no additional runtime cost.

The contribution of this thesis is an improved representation of occluded objects as an extension to the Probabilistic Multiple Hypothesis Anchoring (PMHA) framework from [11]. By assuming that a static map of the environment is known as well as taking previously detected objects into account, areas can be determined in which an object will be occluded. If an object is no longer detected nearby such an area, its position is represented by a distribution over the

[1] `https://athome.robocup.org/`

occluded area. Hereby, the state of the object is more realistically represented. It will be shown that the PMHA framework is better capable of object tracking due to an increased representation of the world by including this extension. In future work, this could be used to limit the growth of the hypothesis tree.

The remainder of this thesis is organized as follows. In Chapter 2 the challenge of people tracking in the scope of the RoboCup@Home competition is further elaborated. In Chapter 3 the related work is discussed and the contributions of this thesis are summarized. Chapter 4 introduces the probabilistic multiple hypothesis framework followed by a detailed description of the extension to this framework in Chapter 5. Several experiments and their results are discussed in Chapter 6 after which a conclusion is given and future work is discussed in Chapter 7.

Chapter 2

Problem Description

The RoboCup@Home competition was initiated 'to develop service and assistive robot technology with high relevance for future personal domestic applications'. In 2011 Tech United Eindhoven[1] joined this competition. Since 2017, the team is competing in the Domestic Standard Platform League (DSPL) [12]. In the DSPL all teams use the same robot platform which is the Human Support Robot (HSR) made by Toyota Motor Cooperation[2]. With this robot, which was named HERO, Tech United became world champion in 2019.

The HSR robot, shown in Figure 2.1, is a compact mobile robot intended to assist elderly and disabled people at home. It can move around the house and is capable to interact with people via its display and speakers. Furthermore, it can manipulate objects using its arm and gripper. The HSR contains a variety of sensors, including a laser scanner and an RGB-D camera. The exact specifications as well as the onboard computation power are given in Table 2.1. Since Tech United is competing in the DSPL this configuration cannot be changed.

In literature, the problem of people tracking is often split up into two sub-problems: 1) detection and 2) Multiple Target Tracking (MTT). The latter is a combination of data association and model-based object tracking. Currently, HERO is only capable of detecting people using OpenPose and RGB-D data. OpenPose is an open-source system for the detection of 2D poses of multiple people in real-time [13]. The faces of detected people are investigated using OpenFace [14]. If a face is recognized in the database it is labeled. For each labeled person, additional properties such as age, gender, and shirt color are identified. Furthermore, a 3D skeleton representation is constructed [12].

For now, it will be assumed that the problem of people detection is solved (Assumption I). The focus of this thesis is to establish an MTT framework that is capable of tracking multiple people real-time in indoor environments for reasons such as collision avoidance, social aware navigation, and person following.

[1] `http://www.techunited.nl`
[2] `https://robots.ieee.org/robots/hsr/`

Table 2.1: HSR specifications

Description	**Configuration**
Laser scanner	UST-20LX
RGB-D camera	Xtion PRO LIVE
CPU	4th Gen Intel Core i7
GPU	NVIDIA Jetson TK1
RAM	16GB
SSD	256GB

Figure 2.1: HERO

In summary, the MTT framework should meet the following requirements which were formulated by [11]:

1. *Real-time execution*: The framework has to operate on a mobile robot such as HERO with limited computation power and battery life. Therefore, the framework should be computationally efficient.
2. *Proper data association*: To correctly update people's attributes such as positions and poses, measurements need to be associated to the corresponding person. However, there are a lot of uncertainties present making this a non-trivial task. Sensor measurements generally contain noise, false positives, and false negatives. Furthermore, similar people might cause ambiguous situations when positioned close to each other. Therefore, a data association algorithm is needed to solve this problem.
3. *Model-based object tracking*: For a variety of tasks like socially aware navigation and people following it is desired to represent the current state of a person, to predict future movements, and to keep track of identities while moving. Several models exist of which an overview is provided in [15]. One of the most significant challenges in an indoor environment is the high occlusion probability. Obstacles such as furniture could temporarily occlude targets. The MTT framework should be able to deal with this naturally.
4. *Appropriate anchoring*: Persons are semantically rich objects. For purposes such as re-identification as well as for more reliable data association, it is crucial to establish a link between these semantically rich objects and measurements [16]. This is called anchoring. For example, the fact that Person 1 wears a red shirt and Person 2 a green one could improve data association if position measurements correspond to a color.

Chapter 3

Related Work

For several decades the problem of multiple target tracking has extensively been studied and various methods were developed over the years. In this chapter, the most relevant work is discussed regarding the problem that was formulated in the previous chapters.

3.1 Data association

One of the most trivial techniques for data association is the Nearest Neighbour (NN) approach [4, 9, 17–19]. For all objects, either the Euclidean or Mahalanobis distance between the object and the measurements is usually computed. For each object, the measurement with the smallest distance is sequentially associated. A measurement can only be associated once. In [20] a Global NN (GNN) approach is used. The difference with the regular NN approach is that the global distance between objects and measurements is minimized at once. The advantages of NN approaches are a straightforward implementation and a low computational cost. A serious limitation is that hard decisions are being made: once a measurement is associated with an object, all the other information is deleted. Therefore, wrong associations in the past cannot be corrected anymore.

A second technique is the Joint Probabilistic Data Association (JPDA) approach [1, 21, 22]. It is a further development of the Probabilistic Data Association (PDA) approach enabling the tracking of multiple targets. Instead of associating a single measurement to a track, an attribute is updated by a weighted average of all the measurements within a certain validation gate. The weight is determined by the probability of associating the measurement to the track. In contrast to the NN approach, a soft decision is being made by including multiple measurements. Thereby, JPDA is less sensitive to clutter. However, it is computationally expensive and the processing time grows exponentially with the number of tracks and measurements. Furthermore, JPDA suffers from track bias and coalescence [23]. To overcome these problems, several variants were developed such as cheap JPDA, sub-optimal JPDA, and NN-JPDA.

Another, more sophisticated technique is the Multiple Hypothesis Tracker (MHT) approach [7, 24–27]. Here, all statistically feasible representations of the world state are considered in parallel. A feasible representation is called a hypothesis and contains one possible set of assignments (i.e. associations) and interpretations of measurements and objects. Measurements are assumed to originate from either previously detected objects, new objects, or false alarms; objects can be interpreted as detected, non-detected, deleted, or occluded. For each hypothesis, a probability is computed based on Bayesian rules. The hypothesis with the highest probability is assumed to correctly represent the world state. By considering multiple hypotheses over time, previous state representations can be corrected based on new measurements. The disadvantage of MHT is its high computational cost which grows exponentially with the number of measurements and objects. Moreover, switching world states could occur because the final decision is postponed.

The NN, JPDA, and MHT approaches all meet the first three requirements formulated in Chapter 2 and have the potential to meet the fourth requirement as well. However, their optimality, implementation, and tracking performance differ. To start with, NN and JPDA can be derived from MHT and are therefore sub-optimal in nature [7, 11]. Furthermore, in contrast to NN and JPDA where track initiation and deletion are based on heuristic rules, MHT handles the whole life cycle of objects. Moreover, in [6, 8, 9, 23, 28] their tracking performance is compared. For simple scenarios (i.e. a single, slowly maneuvering target without the presence of clutter) the performance of the three techniques is similar. Introducing clutter decreases the functioning of the NN approach significantly without a proper initiation and deletion logic, whereas MHT and JPDA show better performance in cluttered environments but at a high computational cost. MHT showed the overall best performance, but the difference with NN disappears in more complex environments (i.e. multiple, highly dynamic targets in the presence of clutter) when the MHT has to run real-time with limited computation power [6, 8].

The fact that real-time MHT applications are superfluous in simple scenarios and have no added value in too complex scenarios was also acknowledged by [7]. Due to the constrained time and computation power, the generation, maintenance, and evaluation of hypotheses are impeded to an extent that its optimality vanishes. Therefore, the growth of the hypothesis tree should effectively be managed. Most current techniques focus on pruning [7]. However, in [10] it is shown that limiting the generation of hypotheses instead of pruning generated hypotheses might be the way to go. The authors used a local grid map to include place-dependent target behavior. This replacement of a simplistic assumption on which MHT is based led to improved tracking results at no additional runtime cost.

3.2 Model-based Object Tracking

To track and propagate object states, motion models are required. Several approaches exist of which an overview is provided in [15]. A physics-based approach is commonly used in literature where the dynamic behavior of an object is modeled as an explicit transition function (e.g. a Constant Velocity (CV) model). Sometimes, multiple models are used in parallel to cover different aspects of the dynamic behavior using an Interacting Multiple Model algorithm [6, 9]. Since physics-based models are always approximations, state observers such as Kalman Filters (KFs) and Particle Filters (PFs) are used to compensate for this [4, 29, 30]. In summary, KFs assume linear (KFs), locally linear (extended KFs), and non-linear (unscented KFs) behavior as well as Gaussian distributions. The computational cost increases respectively. PFs can handle non-linearity and non-Gaussian distributions but at a high computational cost.

In domestic environments, humans are often (partially) occluded by obstacles or other humans. However, observability is often assumed in MTT problems [31]. Without a model that sufficiently represents the object's state during occlusion, MTT techniques struggle to handle occlusions because the uncertainty of the state increases over time to a point where association becomes unlikely. In MHT implementations based on [25], the object is then simply deleted and a new object is created when the person reappears. In [11] the object is represented by a Gaussian distribution located at the point where it was detected for the last time. If the object reappears probabilistically too far away, a new object is created as well.

Several attempts were made to handle occlusions more naturally. In [26] an explicit label for occlusions was incorporated in the MHT framework from [25] to track legs. Once two legs are associated with a single person, the occlusion probability of both legs is adapted to deal with their frequent occlusions. Hereby, the deletion of leg tracks was significantly reduced. However, after a while, the probability of deletion becomes higher than the probability of occlusion. As a result, this method only works for short-lived occlusions.

In [32] the authors solved this by introducing a map for the sensor detection probability. For an object that is likely to be occluded by another object, the probability of detection is decreased. Hereby, the probability of existence does not decrease and, as a result, the object is not deleted. However, the motion model that represented the dynamic behavior before the occlusion event is simply propagated. This does not always cover the real object behavior. Furthermore, the uncertainty about the object's state could still increase to an extent where it is not associated anymore.

A first attempt to solve the latter was made in [30]. The authors limited the propagation window to a certain angle using so-called discrete points of maximum uncertainty. Hereby, the increase in uncertainty was controlled which improved tracking results during short-lived occlusions. However, after 5 seconds of non-detection, the object was still deleted. The same authors further developed their method in [31] by including the fact that a person cannot disappear into nowhere (i.e. deletion does not make sense). Instead of propagating the object state during occlusion, an algorithm was developed that creates a region of uncertainty along the line of reappearance based on the object's direction. However, in both methods it is assumed that a person continuously walks in around the same direction. If a person would turn behind the object, it would not be tracked correctly.

3.3 Contributions

Based on the related work, the MHT approach is preferred since it is optimal compared to NN and JPDA, handles the whole life cycle of an object in a probabilistic way, and shows the overall best performance, partially due to its ability to resolve ambiguous situations over time. In order to meet the fourth requirement, the MHT has to be combined with an anchoring framework. This was done in [11] resulting in a Probabilistic Multiple Hypothesis Anchoring (PMHA) algorithm which will be further elaborated in Chapter 4.

However, still many challenges remain of which two are addressed in this thesis. In the first place, the exponential growth of hypotheses could eliminate the added value of the computational costly MHT. Secondly, the handling of occlusions is only solved under certain assumptions such as continuous walking behavior. In this thesis, a method is proposed in which the latter is solved without the assumption of continuously moving objects. The state of an occluded object is represented by a distribution over the area that is out of sight. It is believed that this improved representation of occluded object states could be used in future work to limit the growth of the hypothesis tree.

The contributions of this thesis are:

- An improved representation of occluded targets as an extension to the PMHA framework from [11],
- An experimental validation of the extension's added value for various simulated scenarios,
- A ROS-based framework to efficiently generate various simulation rooms and trajectories.

Chapter 4

Probabilistic Multiple Hypothesis Anchoring Algorithm

In this chapter, an overview of the PMHA algorithm is given that was originally developed in [11]. To start with, physical objects with corresponding attributes are stored in a symbol system (i.e. a database). Each object is represented by a unique individual symbol $\ell = \{\iota_1, \iota_2, \ldots\}$ and a set of static, predicate symbols $\mathcal{P}$. For example, a person could have the individual symbol $\iota_1 =$ `Person-A` and the set of predicate symbols $\mathcal{P} = \{$ `human, face-A, tall`$\}$. This semantically rich information is used in the world model. During operation, a robot measures certain attributes like positions, faces, and dimensions. These measurements are mapped to predicate symbols using predicate grounding relations (e.g. `tall` is a body length > 1.9 m). Hereby, an object can be linked to an individual symbol.

In the world model, an object is represented by an anchor α. Some of the anchor's attributes are updated with measurements. Let $Z(k) = \{\boldsymbol{z}_i^k\}_{i=1}^{n_{\text{meas}}^k}$ be the set of n_{meas}^k measurements at discrete time step k and $Z^k = \{Z(j)\}_{j=1}^k$ be the set of all measurements up to time k. An anchor is then represented as:

$$\alpha_a = \left(\iota, \boldsymbol{z}_i^k, \mathcal{M}_a^k\right), \tag{4.1}$$

where $a \in [1, n_{\text{obj},h}]$ is the anchor index, ι is an individual symbol, and $\boldsymbol{z}_i^k$ is a measurement linked to this symbol with index $i \in [1, n_{\text{meas}}]$ at time k. $\mathcal{M}_a^k$ is the set of behavior models given by:

$$\mathcal{M}_a^k = \begin{cases} p\left(M_{a,1}^k\right) & : \quad M_{a,1}^k \\ \vdots & \quad\quad \vdots \\ p\left(M_{a,n_{\text{mod},a}}^k\right) & : \quad M_{a,n_{\text{moda}}}^k \end{cases} \tag{4.2}$$

Here, $M_{a,m}$ is a behavior model with index m, $p\left(M_{a,m}\right)$ is the Probability Mass Function (PMF) over the behavior models evaluated at $M_{a,m}^k$, and $n_{\text{mod},a}$ is the number of behavior models in anchor α. Per object class (e.g. `person` or `cup`) the behavior models can be specified.

In addition, an estimate for an object's attribute is given by a Posterior Density Function (PDF) $\gamma_{a,m}^k$ based on the corresponding behavior model. For example, in case of the attribute 'position' and a KF with a CV model as a behavior model, the PDF $\gamma_{a,m}^k$ is the KF's Gaussian distribution. The final attribute's estimate Γ_a^k is computed as the convex sum over the estimates of all behavior models in $\mathcal{M}_a^k$:

$$\Gamma_a^k = \sum_{m=1}^{n_{\text{mod},a}} p\left(M_{a,m}^k\right) \gamma_{a,m}^k \tag{4.3}$$

In order for anchors to update their attributes, measurements have to be associated accordingly. This is done with an MHT that is based on [24]. It is assumed that measurements originate from i) a new anchor, ii) a previously detected (i.e. an existing) anchor, and iii) a false detection (i.e. clutter). Furthermore, an anchor is labeled as detected or non-detected. It should be noted that the deletion of anchors is not handled by the PMHA algorithm in contrast to [25, 26].

For a new set of measurements $Z(k)$ at time step k, the current anchors are propagated first according to their behavior models. Thereafter, the measurements are associated. Each feasible set of associations between measurements and new objects, existing objects, or clutter is represented by an assignment set θ^k and forms a possible representation (i.e. hypothesis) of the world. A hypothesis Θ_h^k with $h \in [1, n_{\text{hyp}}]$ contains a set of anchors with corresponding attributes and is preceded by a parent hypothesis $\Theta_{p(h)}^{k-1}$. Hence, θ^k maps $\left\{\Theta_{p(h)}^{k-1}, Z(k)\right\}$ to Θ_h^k. The posterior probability of a hypothesis Θ_h^k given all measurements Z^k is computed via Bayes' law:

$$p\left(\Theta_h^k|Z^k\right) = \frac{p\left(Z(k)|\Theta_h^k, Z^{k-1}\right) p\left(\theta_h^k|\Theta_{p(h)}^{k-1}, Z^{k-1}\right) p\left(\Theta_{p(h)}^{k-1}|Z^{k-1}\right)}{p\left(Z(k)|Z^{k-1}\right)} \tag{4.4}$$

where $p\left(Z(k)|\Theta_h^k, Z^{k-1}\right)$ is the likelihood (i.e. the probability of the measurements given the hypothesis), $p\left(\theta_h^k|\Theta_{p(h)}^{k-1}, Z^{k-1}\right)$ is the prior (i.e. the probability of θ_h^k given its parent hypothesis and all measurements up to time $k-1$), $p\left(\Theta_{p(h)}^{k-1}|Z^{k-1}\right)$ is the posterior probability of the parent hypothesis, and $p\left(Z(k)|Z^{k-1}\right)$ is the normalization term. The likelihood is computed as:

$$p\left(Z(k)|\Theta_h^k, Z^{k-1}\right) = \prod_{i=1}^{n_{\text{meas}}} p\left(\boldsymbol{z}_i^k|\alpha_{h,a_i}^k, Z^{k-1}\right) = V^{-n_{N,h}^k - n_{F,h}^k} \cdot \prod_{i=1}^{n_{\text{meas}}} p\left(\boldsymbol{z}_i^k|\Gamma_a^k\right)^{\delta_i} \tag{4.5}$$

$$p\left(\boldsymbol{z}_i^k|\Gamma_a^k\right) = \prod_{m=1}^{n_{\text{mod},a}} p\left(M_{h,a_i,m}^k\right) p\left(\boldsymbol{z}_i^k|\gamma_{h,a_i,m}^k\right) \tag{4.6}$$

where a_i is the index of the anchor to which measurement $\boldsymbol{z}_i^k$ is associated, V is the observable volume in which the detection of new anchors and false positives is assumed to be uniform, $n_{N,h}^k$ and $n_{F,h}^k$ are the number of new anchors and false positives respectively, and δ_i is an indicator variable that equals 1 if measurement $\boldsymbol{z}_i^k$ is associated with an anchor and 0 otherwise. The prior is given by:

$$p\left(\theta^k|\Theta_{p(h)}^{k-1}, Z^{k-1}\right) = \frac{n_{N,h}^k! n_{F,h}^k!}{n_{\text{meas}}^k!} p_N\left(n_{N,h}^k\right) p_F\left(n_{F,h}^k\right) \cdot \prod_{a=1}^{n_{\text{obj},h}} \left(p\left(D_{h,a}^k\right)\right)^{\delta_a} \left(1 - p\left(D_{h,a}^k\right)\right)^{1-\delta_a} \tag{4.7}$$

$$p\left(D_{h,a}^k\right) = p\left(D_{h,a}^k|V_{h,a}^k\right) \sum_{m=1}^{n_{\text{mod},a}} p\left(V_{h,a}^k|M_{h,a,m}^k\right) p\left(M_{h,a,m}^k\right) \tag{4.8}$$

where $p_N\left(n_{N,h}^k\right)$ and $p_F\left(n_{F,h}^k\right)$ are the prior PMFs for the number of new anchors and false positives respectively and $p\left(D_{h,a}^k\right)$ is the probability of detection assumed that only visible objects are detected.

In order to limit the growth of the hypothesis tree, two pruning strategies are implemented. In the first place, the posterior probability of a hypothesis has to be above a user-defined threshold to prune really unrealistic hypotheses. Secondly, an algorithm is used to generate only the n_{hyp} most probable hypotheses. It should be noted that this is an approximation due to which it cannot be guaranteed that the correct hypothesis is maintained. To conclude, the most probable hypothesis is believed to correctly represent the world.

Chapter 5

Occlusion Handling Extension

In Chapter 4, the PMHA algorithm is introduced that was developed in [11]. In this chapter, the extension for this algorithm is elaborated. The goal of the extension is to improve the handling of occlusions. In the original PMHA package, positions are tracked with a CV model and a KF. To limit the expansion of the uncertainty, a Gaussian distribution with a fixed covariance is located at the last detected point when no new measurements are received for `kalman_timeout` seconds. As a result, a new anchor is created when the object reappears probabilistically too far away (see Figure 5.1).

In the RoboCup@Home competition, it is allowed to create a static map of the indoor environment (Assumption II). Based on this map a robot could determine areas where objects cannot be detected (i.e. are occluded) with respect to its current position. In the extension, this information is used to adjust the behavior model of an anchor. For reasons of simplicity, it is assumed that the occluded area is fully enclosed by a detectable area (Assumption III) and that objects cannot leave the occluded area except through a detectable area (e.g. no doors or stairs in the occluded area) (Assumption IV). With these assumptions, the position of an object in an occluded area can be represented by a uniform distribution over the occluded area (see Figure 5.2). Hereby, the uncertainty in position is time-invariant and the reappearing object can always be correctly associated. Furthermore, this method does not require continuous walking behavior in contrast to [30, 31]. In Algorithm 1 the pseudo-code of the extension's implementation in the propagating step of the behavior model is given.

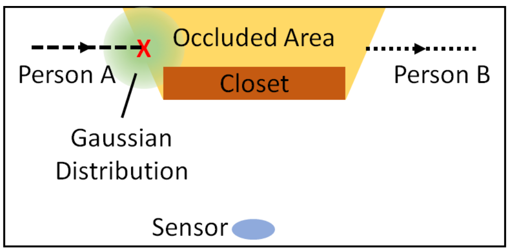

Figure 5.1: PMHA without extension: person A walks behind a closet. At the red cross, the sensor does not detect person A anymore. A Gaussian distribution located at the point of the last detection represents the current position. After reappearing, a new person B is created.

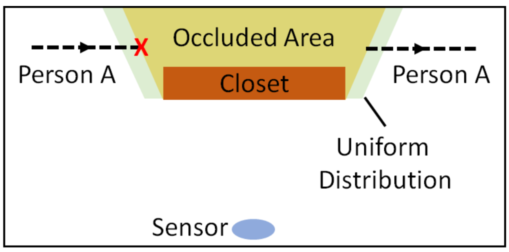

Figure 5.2: PMHA with extension: person A walks behind a closet. At the red cross, the sensor does not detect person A anymore. A uniform distribution over the occluded area represents the current position. After reappearing, the detections are associated to person A again.

Algorithm 1: Propagation step in the behavior model for the extended PMHA

```
Input  : time
Output: propagated_position, propagated_pdf
last_detected_position = mean(kalman_filter);
if (time - t_last_update) > kalman_timeout then
    if last_detected_position ∈ occluded_area then
        propagated_position = mean(occluded_area);
        propagated_pdf = Uniform(occluded_area);
        return;
    else
        propagated_position = fixedState(last_detected_position);
        propagated_pdf = Gaussian(fixed_cov);
        return;
    end
else
    propagated_position = propagate(kalman_filter);
    propagated_pdf = Gaussian(kalman_filter);
end
```

Chapter 6

Experimental Validation

In the previous chapters, both the PMHA algorithm and the occlusion handling extension that was developed in this thesis are discussed. In this chapter, the functioning and added value of the extension is investigated for several scenarios.

6.1 Experimental Set-up

To validate the functioning and added value of the occlusion handling extension, a comparison is made between the PMHA algorithm with and without the extension for several scenarios. These scenarios were simulated using the Robot Operating System (ROS) framework on a PC for which the specifications are given in Table 6.1. To start with, the occlusion handling extension has been incorporated in the PMHA ROS package[1] as a separate behavior model. In addition, a separate package was created to generate rooms and to simulate position detections of objects. A trajectory is defined as a continuous line between certain points. An object follows this line with a constant velocity of either 0 or 0.5 m/s. At a frequency of 10 Hz, discrete points are sampled from the object's position to which detection noise is added according to a Gaussian distribution $\mathcal{N}(0, 0.001)$. These 'detections' are published to the PMHA package via a ROS node. This setup makes it possible to easily replace the simulated detections by detections from a real sensor. Besides, a set containing the boundaries of the occluded area is published to the PMHA package as well: $\{x_{min}, x_{max}, y_{min}, y_{max}\}$. This information is used by the extended behavior model to determine the occluded area. To conclude, a room is created by defining markers that are published to `rviz` over a separate node. The resulting packages are available on GitHub[2].

[1] `http://wiki.ros.org/wire`
[2] `https://github.com/tue-robotics/bep_tpcw` & `https://github.com/tue-robotics/wire/tree/bep_tpcw`

Table 6.1: Specifications PC used during experiments

Description	**Configuration**
OS	Ubuntu 16.04
CPU	7th Gen Intel Core i7
GPU	Intel HD Graphics 630 NVIDIA Quadro M1200
RAM	8GB
SSD	256GB

6.2 Experiments

The original and extended PMHA package are compared for several simulated scenarios. In this section, these experiments and the resulting world representation for both packages are discussed. A video of the simulations is available on Youtube[3].

6.2.1 Scenario 1: Single, Temporary Occluded Person

In the first scenario, depicted in Figure 6.1, a single person follows a trajectory that is halfway occluded by a closet. The goal is to validate that the occluded person is correctly associated with the track after reappearing. For the original package, this is not the case. As can be seen in Figure 6.2, the person is tracked until the point of occlusion. Then, this person is represented by a Gaussian distribution with a fixed covariance located at the point where the person was last detected. Once the person reappears, it is not associated with the person with the fixed distribution, and a new person is created resulting in a wrong world model. However, for the extended package the reappearing person is correctly associated as shown in Figure 6.3 and, as a result, the world model is in accordance with reality.

6.2.2 Scenario 2: Appearance of a New Person

The fact that the robot cannot detect anything in the occluded area gives room for a scenario where a new person is located in the occluded area beforehand. If a detected person disappears in the occluded area, and this new person appears afterward, the framework should be able to model this correctly. However, based on position measurements only, it is impossible to distinguish these two persons. Therefore, it is assumed that another sensor is available that detects a more person bonded feature (e.g. face or shirt color). For now, a color is used. The color is published to the PMHA package with a detection probability of 10%. In the second scenario, this situation is simulated (Figure 6.4). Despite the low color probability, both the original and extended packages handle this correctly as can be seen in Figures 6.5 and 6.6.

6.2.3 Scenario 3: Two Simultaneously Occluded Persons

In the third scenario, two persons that follow mirrored trajectories are simulated (Figure 6.7). Each person has a unique color (i.e. green and red) for which the detection probability is the same as in scenario 2. The goal is to validate a correct representation of the world in case of a simultaneous occlusion event. In Figure 6.8 it can be seen that for the original package the reappearing persons

[3] `https://youtu.be/6RL978VSAWA`

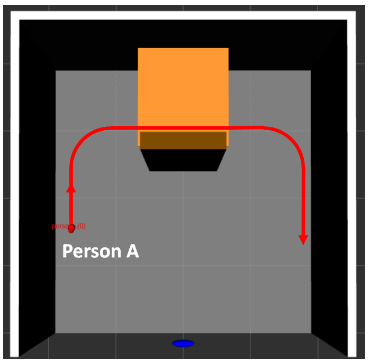

Figure 6.1: Scene-1 ground-truth: Person A walks continuously behind a closet.

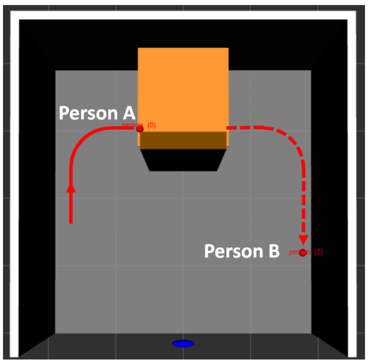

Figure 6.2: Scene-1 result old: Person A is tracked to the point where it is occluded. Person B is created after reappearing. $n_{hyp} = 2$

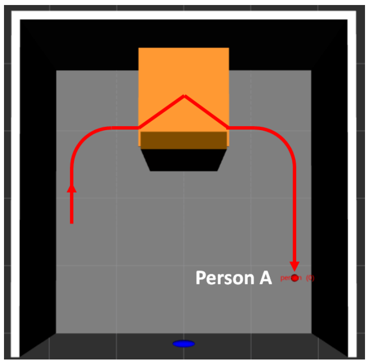

Figure 6.3: Scene-1 result new: Person A is tracked to the point where it is occluded and continued after reappearing. $n_{hyp} = 2$

are associated with the wrong track, even despite the fact that the detected color does not match. This is a consequence of the low detection probability. However, in the extended package, the association is done correctly as can be seen in Figure 6.9. Furthermore, it requires one hypothesis less compared to the original package due to its improved world representation.

6.2.4 Scenario 4: Person Occluded by Another Person

In contrast to the previous three scenarios, persons can also be occluded by dynamic objects such as other persons. The handling of dynamic occlusions is investigated in this fourth and last scenario (Figure 6.10). A red person follows a U-shape trajectory and a green person walks behind the red person. The detection probability is the same as in scenario 2. At a certain point, the green person is occluded by the red person. After a while, the green person reappears again. For the original package, the person is not associated correctly. At the point of occlusion, the person's position is fixed and after reappearing, a new person is created (Figure 6.11). The extended package however has the correct most probable hypothesis since the occluded person is associated correctly after reappearing (Figure 6.12).

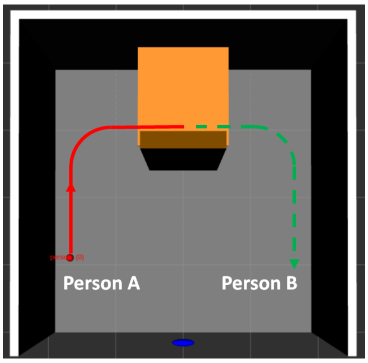

Figure 6.4: Scene-2 ground-truth: Person A walks continuously behind a closet. Thereafter, Person B appears.

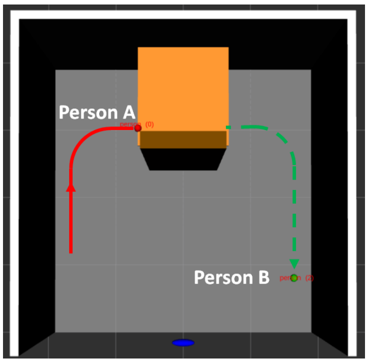

Figure 6.5: Scene-2 result old: Person A is tracked to the point where it is occluded. At appearance, person B is created. $n_{hyp} = 2$

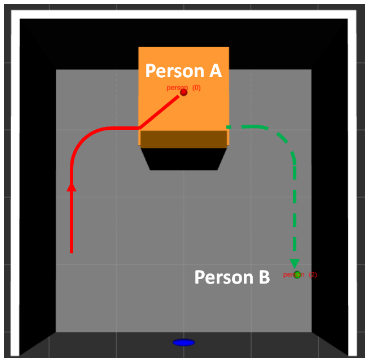

Figure 6.6: Scene-2 result new: Person A is tracked to the point where it is occluded. At appearance, person B is created. $n_{hyp} = 2$

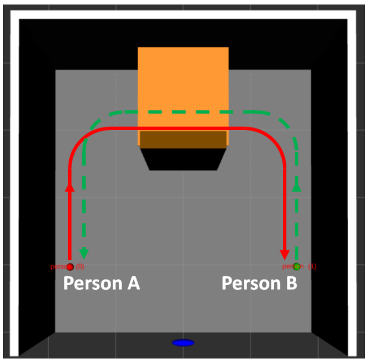

Figure 6.7: Scene-3 ground-truth: Person A and B walk continuously along a mirrored trajectory that is halfway occluded by a closet.

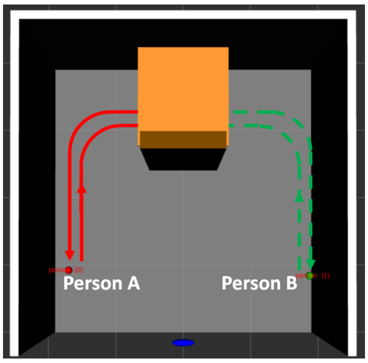

Figure 6.8: Scene-3 result old: Both persons are tracked to the point of occlusion and are associated to the wrong track after reappearing. $n_{hyp} = 5$

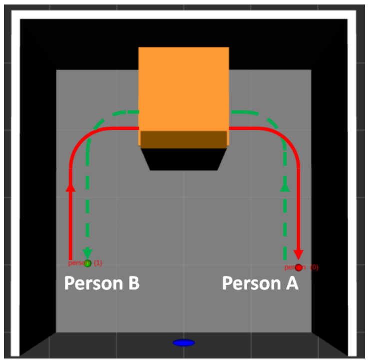

Figure 6.9: Scene-3 result new: Both persons are tracked to the point of occlusion and are associated to the right track after reappearing. $n_{hyp} = 4$

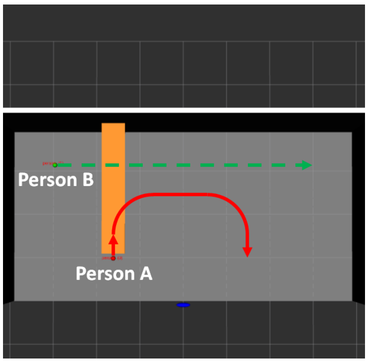

Figure 6.10: Scene-4 ground-truth: Person A walks continuously along the red line. Person B walks in a straight line and is temporary occluded by person A.

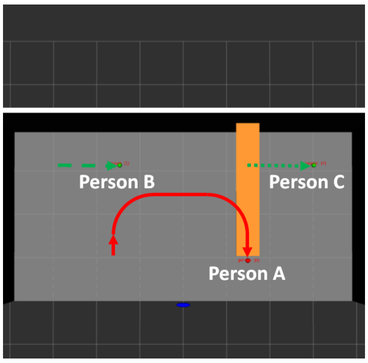

Figure 6.11: Scene-4 result old: After occlusion, the position of person B is fixed. A new person C is created after reappearing. $n_{hyp} = 3$

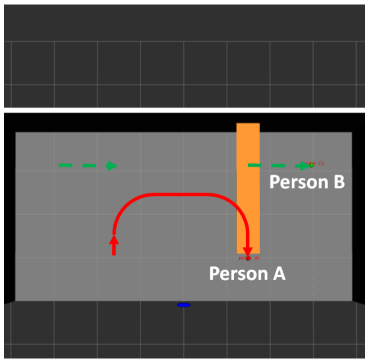

Figure 6.12: Scene-4 result new: During occlusion, the position of person B is maintained in the occluded area of person A. After reappearing person B is associated again. $n_{hyp} = 3$

Chapter 7

Conclusions

In this thesis, an extension to the PMHA algorithm from [11] is proposed to improve the handling of occlusions. This is achieved by adapting the behavior model of the anchors. Based on a static map of the environment, areas are determined in which an object cannot be detected by the robot relative to its current position. These areas are labeled as occluded areas. In case the track of an object is lost nearby an occluded area for a certain amount of time, the Kalman filter is replaced by a uniform distribution over the occluded area. Hereby, the uncertainty in position is time-invariant and association after reappearing is possible along the edges of the occluded area.

The extension's added value was validated in several experiments. For all simulated scenarios, the extended PMHA package provided a correct representation of the real world in contrast to the original package. For the latter, positions were not correctly propagated, and, as a result, wrong associations were made or new anchors were created. Moreover, it is shown that the extended version has lower requirements for the detection probability of a human-specific attribute. This could be advantageous because these attributes are generally hard to detect.

Future work is required to increase the applicability of the developed extension. To start with, its functioning has only been validated for simulated environments. However, this approximation of the real world could result in different tracking behavior. Therefore, it is recommended to conduct more elaborate experiments with real sensor data. Secondly, some of the assumptions behind the current extension will often be violated. More research is needed to deal with multiple occluded areas that are height dependent, interfere with each other, are temporary out of sight, and contain exits such as doors and stairs.

Furthermore, a time-invariant, uniform distribution is currently used to represent a position in an occluded area. Developing a more sophisticated, time-dependent distribution could further increase the association performance. Besides, right now an if-statement is used to switch between the two behavior models. Using a multiple model approach to probabilistically determine the correct behavior model could be further investigated.

To conclude, research into a more sophisticated control for the growth of the hypothesis tree would be highly interesting based on this improved representation of the world during occlusions. When combined with the work from [10] one might be able to distinguish between ambiguous and trivial situations and change the generation and pruning of hypotheses accordingly.

Acknowledgements

The author would like to thank Josja Geijsberts, Lars Janssen, René van de Molengraft, Jos Elfring, Laura de Jong, Gosse Bijlenga, Matthijs van der Burgh, and Pepijn Griffioen for their gracious help during the realization of this thesis.

Bibliography

[1] G. Bijlenga, "Semantic Classification and Linking of 2D-LIDAR Filtered Low-Level Features," 2019. 1, 5

[2] L. Beyer, A. Hermans, T. Linder, K. O. Arras, and B. Leibe, "Deep Person Detection in 2D Range Data," 2018. 1

[3] K. O. Arras, B. Lau, S. Grzonka, M. Luber, O. M. Mozos, D. Meyer-Delius, and W. Burgard, "Range-based people detection and tracking for socially enabled service robots," *Springer Tracts in Advanced Robotics*, vol. 76, no. STAR, pp. 235–280, 2012. 1

[4] N. Bellotto and H. Hu, "Computationally efficient solutions for tracking people with a mobile robot: An experimental evaluation of Bayesian filters," *Autonomous Robots*, vol. 28, pp. 425–438, 5 2010. 1, 5, 6

[5] Y. Bar-Shalom and X.-R. Li, "Multitarget-Multisensor Tracking: Principles and Techniques," *IEEE Aerospace and Electronic Systems Magazine*, vol. 11, no. 2, p. 41, 1995. 1

[6] M. De Feo, A. Graziano, R. Miglioli, and A. Farina, "IMMJPDA versus MHT and Kaiman filter with NN correlation: Performance comparison," *IEE Proceedings: Radar, Sonar and Navigation*, vol. 144, no. 2, pp. 49–56, 1997. 1, 6

[7] C. Y. Chong, S. Mori, and D. B. Reid, "Forty Years of Multiple Hypothesis Tracking - A Review of Key Developments," *2018 21st International Conference on Information Fusion, FUSION 2018*, pp. 452–459, 2018. 1, 5, 6

[8] T. Linder, S. Breuers, B. Leibe, and K. O. Arras, "On multi-modal people tracking from mobile platforms in very crowded and dynamic environments," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2016-June, pp. 5512–5519, 2016. 1, 6

[9] T. Linder, F. Girrbach, and K. O. Arras, "Towards a Robust People Tracking Framework for Service Robots in Crowded, Dynamic Environments," *International Conference on Intelligent Robots and Systems (workshop) - IROS*, 2015. 1, 5, 6

[10] M. Luber, G. Diego Tipaldi, and K. O. Arras, "Place-dependent people tracking," *International Journal of Robotics Research*, vol. 30, no. 3, pp. 280–293, 2011. 1, 6, 16

[11] J. Elfring, S. Van Den Dries, M. J. Van De Molengraft, and M. Steinbuch, "Semantic world modeling using probabilistic multiple hypothesis anchoring," *Robotics and Autonomous Systems*, vol. 61, no. 2, pp. 95–105, 2013. 1, 4, 6, 7, 8, 10, 16

[12] M. v. d. Burgh, J. Lunenburg, R. Appeldoorn, R. Wijnands, T. Clephas, M. Baeten, L. v. Beek, R. Ottervanger, H. v. Rooy, and M. v. d. Molengraft, "Tech United Eindhoven @Home 2020 Team Description Paper," *RoboCup@Home*, pp. 1–9, 2020. 3

[13] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, no. Xxx, pp. 1302–1310, 2017. 3

[14] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications," tech. rep., 2016. 3

[15] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human Motion Trajectory Prediction: A Survey," 5 2019. 4, 6

[16] M. J. Islam, J. Hong, and J. Sattar, *Person-following by autonomous robots: A categorical overview*, vol. 38. 2019. 4

[17] C. Dondrup, N. Bellotto, F. Jovan, and M. Hanheide, "Real-time multisensor people tracking for human-robot spatial interaction," *Workshop on Machine Learning for Social Robotics at International Conference on Robotics and Automation (ICRA)*, 2015. 5

[18] N. Bellotto and H. Hu, "Vision and laser data fusion for tracking people with a mobile robot," in *2006 IEEE International Conference on Robotics and Biomimetics, ROBIO 2006*, pp. 7–12, 2006.

[19] F. Beck and M. Bader, "Map Based Human Motion Prediction for People Tracking," pp. 1–7, 2020. 5

[20] A. Leigh, J. Pineau, N. Olmedo, and H. Zhang, "Person tracking and following with 2D laser scanners," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2015-June, no. June, pp. 726–733, 2015. 5

[21] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Sonar Tracking of Multiple Targets Using Joint Probabilistic Data Association," *IEEE Journal of Oceanic Engineering*, vol. 8, no. 3, pp. 173–184, 1983. 5

[22] D. Schulz, W. Burgard, D. Fox, and A. B. Cremers, "People tracking with mobile robots using sample-based joint probabilistic data association filters," *International Journal of Robotics Research*, vol. 22, no. 2, pp. 99–116, 2003. 5

[23] Z. Hu, H. Leung, and M. Blanchette, "Evaluation of data association techniques in a real multitarget radar tracking environment," *Signal and Data Processing of Small Targets 1995*, vol. 2561, no. September 1995, pp. 509–518, 1995. 5, 6

[24] D. B. Reid, "An Algorithm for Tracking Multiple Targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979. 5, 9

[25] 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 138–150, 2 1996. 6, 9

[26] 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 *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 1710–1715, IEEE Xplore, 2008. 6, 9

[27] S. Müller, T. Wengefeld, T. Q. Trinh, D. Aganian, M. Eisenbach, and H. M. Gross, "A multimodal person perception framework for socially interactive mobile service robots," *Sensors (Switzerland)*, vol. 20, no. 3, 2020. 5

[28] J. Smith, F. Particke, M. Hiller, and J. Thielecke, "Systematic Analysis of the PMBM, PHD, JPDA and GNN Multi-Target Tracking Filters," *FUSION 2019 - 22nd International Conference on Information Fusion*, 2019. 6

[29] J. Fernández-Villaverde, "Kalman and particle filtering," in *Macroeconometrics and Time Series Analysis*, pp. 151–157, 2010. 6

[30] H. M. Nemati, "Detection and Tracking of People from Laser Range Data," 2010. 6, 7, 10

[31] H. M. Nemati, S. G. Shahbandi, and B. Åstrand, "Human tracking in occlusion based on reappearance event estimation," *ICINCO 2016 - Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics*, vol. 2, pp. 505–512, 2016. 6, 7, 10

[32] L. Lamard, R. Chapuis, and J. P. Boyer, "Dealing with occlusions with multi targets tracking algorithms for the real road context," *IEEE Intelligent Vehicles Symposium, Proceedings*, pp. 371–376, 2012. 7