Semantic world modeling using probabilistic multiple hypothesis anchoring

J. Elfring *, S. van den Dries, M.J.G. van de Molengraft, M. Steinbuch

Department of Mechanical Engineering, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands

A R T I C L E   I N F O

Article history:
Received 28 June 2012
Received in revised form
29 October 2012
Accepted 9 November 2012
Available online 1 December 2012

Keywords:
Data association
Anchoring
Probabilistic multiple hypothesis anchoring
World modeling
Multiple model tracking

A B S T R A C T

In order to successfully perform typical household tasks such as manipulation or navigation, domestic robots need an accurate description of the world they are operating in. Creating and maintaining such a description, in this work referred to as world model, is a non-trivial task in a domestic environment that typically has a high number of objects, and is unstructured and dynamically changing. This work introduces probabilistic multiple hypothesis anchoring to create and maintain a semantically rich world model using probabilistic anchoring. Multiple hypothesis tracking-based data association is included to be able to deal with ambiguous scenarios. Multiple model tracking is included to be able to easily incorporate different kinds of prior knowledge.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Domestic robots are typically confronted with a complex, dynamically changing and unstructured environment. In such environments, robots must be able to perform a wide variety of tasks, e.g., safe navigation, object manipulation, human–robot interaction, including many different objects. An important prerequisite for successfully accomplishing these tasks is the availability of an accurate description of the part of the environment that is relevant for completing the task. In this work, such an environmental description will be referred to as world model and this work focuses on creating one. With the availability of a world model, a task reasoner can exploit the world description during task planning or execution.

A world model usually contains 3D-positions of uniquely labeled objects [1]. In dynamic scenarios, dynamic object properties such as velocities are included as well [2,3] but depending on the task and the objects that are involved, any object attribute can be added, e.g., color, shape, size. The robot perceives the world using a perceptual system possibly containing multiple sensors such as cameras and laser scanners. The perceptual system generates measurements, i.e., time-stamped features, that can be anything from colored blobs in a camera image [4] to object shapes [5]. These measurements will be the input for the algorithm that constructs the world model. Four requirements for a world modeling algorithm are formulated:

1. Appropriate anchoring: The world model must contain semantically rich objects such that a robot can understand human’s instructions. Therefore, the world modeling algorithm has to link measured attributes to semantically rich objects in the world model. Creating and maintaining this link between instances of grounded object classes and measurements is called anchoring [6].

2. Proper data association: Updating the object attributes based on measurements is a non-trivial task due to the uncertainties a robot is confronted with. Sensors are prone to measurement noise and object detection algorithms inevitably miss objects, i.e., false negatives, or generate false detections, i.e., false positives. Associating measurements with the corresponding object in the world model that generated the measurement can be difficult if visually identical objects appear close to each other, i.e., if the robot faces ambiguities. Data association algorithms are developed to deal with the association problem in a probabilistic manner explicitly taking into account the uncertainty [7].

3. Model-based object tracking: Objects in domestic environments change position regularly, e.g., humans or small household items. Despite these movements, we would like to be able to keep track of the object identities. Various types of models are available to predict the object movements. A physics-based model might describe human movements with a constant velocity motion model, whereas a more common sense based model might describe the location of a person as ‘at home in the evening and at work during the day’. Exploiting various kinds of prior knowledge represented by various types of models, allows more effective object tracking.

* Corresponding author.
E-mail address: J.Elfring@tue.nl (J. Elfring).

0921-8890/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2012.11.005
4. Real-time execution: We would like the robot to perform its task in real-time which limits the maximum allowed computation time. The limited computational resources a robot has limit the allowable computational complexity of the world modeling algorithm.

2. Related work

This section discusses literature related to the problem that was introduced in Section 1.

2.1. Anchoring

In the field of artificial intelligence, the first “systematic study of anchoring as a problem per se” was presented in [8] and refined in [6,9–11]. Next to meeting Requirements 1 and 4, these anchoring strategies all allow for the incorporation of tracking (Requirement 3). However, although the anchoring problem cannot be solved without data association as formulated in Requirement 2, data association is not explicitly dealt with and hence the approach is not able to deal with ambiguous situations.

As a possible solution, [12] extended the work of [6] by adding a strategy that collects additional sensory information with the aim of resolving the ambiguous situation, e.g., let the robot drive around the object and detect visual differences that were invisible from the robot’s original position. However, this solution is considered to be incomplete, since this often is impossible or undesirable.

An extension of [6] focusing on cooperative anchoring in multi-robot systems is [13]. Their approach can be used for single robots as well and data association is included. However, their nearest neighbor data association does not meet Requirement 2 in a satisfactory way as will become apparent in the next section.

An alternative anchoring approach is introduced in [14]. The approach in [14], in addition to Requirement 1, also meets Requirement 2, since the use of a Markov logic network enables probabilistic data association. However, object tracking as defined in Requirement 3 is excluded in this work which complicates dynamic scenarios. In addition, the approach is computationally complex already with a very limited number of objects, hence Requirement 4 will not be met in realistic scenarios containing many objects.

One more alternative is introduced by [15]. Even though the term anchoring is not used, their system can be interpreted as an anchoring algorithm. The main advantage of this method over [6,10,12,14] is the incorporation of a Bayesian framework and the usage of a data association algorithm. However, their data association algorithm is rather simple and, therefore, unable to deal with ambiguous situations (Requirement 2).

Based on the existing literature mentioned in this section, it can be concluded that in the context of our requirements, a proper data association algorithm is missing in the current anchoring algorithms. Proper data association means considering all plausible alternatives if a robot faces ambiguous situations. Incorporating an advanced probabilistic data association algorithm, that does not make silent decisions but instead postpones the final decision until the ambiguity is resolved, into an anchoring strategy is one of the main contributions of this work. For that reason we continue with a section describing data association literature that can be combined with the current anchoring approaches.

2.2. Data association

In literature, data association (Requirement 2) and model-based object tracking (Requirement 3) are usually combined. The combined problem is called multiple target tracking (MTT) and this section reviews the most relevant strategies in the context of this work, whereas [5] gives a more complete overview.

The goal of the Probability Hypothesis Density (PHD) Filter [16] is to find a MTT strategy statistically analogous to the linear Kalman filter (KF) in single-target tracking. Due to the required assumption on the lack of ordering, target identities cannot be recorded [5] and, therefore, the PHD filter is not suitable for meeting Requirement 1.

A popular alternative to the PHD filter is the Multiple Hypothesis Tracker (MHT), originally developed in [17]. The key idea underlying the MHT is that all possible explanations from measurements to objects in the world model are considered in parallel. Each explanation is represented by a hypothesis, hence each hypothesis describes a possible state of the world. Each hypothesis gets a probability of being correct and based on this probability, the robot decides what the world looks like. One important advantage of the MHT is that maintaining multiple hypotheses allows correcting previous conclusions based on new evidence, i.e., new measurements. The MHT is both exponential in time and memory such that in practice only a limited number of hypotheses can be maintained. Therefore, unlikely hypotheses are pruned and as a result the correct solution is not guaranteed to be part of the remaining set of hypotheses.

Two noticeable alternatives to the MHT are the Global Nearest Neighbor (GNN) approach [18] and the Joint Probabilistic Data Association Filter (JPDAF) [7]. If, at each time step, the most likely hypothesis in the MHT is believed to represent the current state of the world and all but the most probable hypothesis are pruned at each time step, the MHT and the GNN approach give identical results, whereas the JPDAF results can be obtained from an MHT by representing the world by a probabilistically weighted average of a subset of hypotheses and again prune all other hypotheses. Both the GNN and the JPDAF need extensions such as [19] in order to allow a varying and unknown number of targets, whereas the MHT deals with this in a very natural way.

The GNN approach, the JPDAF and the MHT all meet Requirements 2, 3 and 4, however the MHT is preferred here since (i) contrary to the GNN approach and the JPDAF, decisions in ambiguous situations are postponed until new evidence resolves the uncertainty, making it the “preferred method for solving the data association problem” [20] and (ii) it has been applied successfully in highly dynamic environments [4].

As mentioned before, data association algorithms (Requirement 2) usually incorporate model-based object tracking (Requirement 3). The way this model-based object tracking is performed is independent of the way the data association is performed. One particularly interesting way of model-based object tracking is covered by Multiple model (MM) methods [21], i.e., tracking using multiple motion models in parallel. Even though MM literature focuses on physics-based motion models, we do not see conceptual problems in adding any kind of prior knowledge that can be represented by a mathematical model, e.g., “John arrives at work around 8 am” can conceptually equally well be included in the model set as “humans approximately move according to a constant velocity motion model”.

In typical tracking applications [22,23] the MHT has proven to be successful and combining it with MM methods is straightforward [20]. However, typical tracking applications exclude the anchoring problem. We think that anchoring is crucial in world modeling for domestic robotics and therefore this work will incorporate an MHT in an anchoring strategy such that the world model contains semantically rich objects.

3. Contributions

We believe that a successful world modeling algorithm must fulfill the Requirements 1–4. For that reason we incorporate the data association of the MHT and MM tracking into an anchoring algorithm. On the one hand the MHT solves data
association in a probabilistic manner (Requirement 2). It (i) uses a probabilistic framework that explicitly deals with uncertainties and (ii) facilitates correcting previous conclusions based on new evidence by maintaining multiple world state hypotheses. On the other hand the anchoring algorithm ensures a semantically rich world model (Requirement 1) and the MM object tracking allows exploiting various kinds of prior knowledge during the model-based object tracking (Requirement 3). As a result, a Probabilistic Multiple Hypothesis Anchoring (PMHA) algorithm is proposed. The contributions of this paper are:

- An anchoring algorithm is extended with MHT resulting in a probabilistic anchoring approach with explicit data association and semantically rich objects (Requirements 1 and 2).
- The definition of objects in the world model is extended compared to existing anchoring algorithms allowing the incorporation of generalized MM object tracking with various types of models (Requirement 3).
- Experimental validation of the presented framework in various challenging scenarios (Requirement 4).

The remainder of this paper is organized as follows. In Section 4, we explain how the anchoring using multiple tracking models is performed in this paper. Then in Section 5 we combine anchoring with multiple hypothesis tracking resulting in a probabilistic multiple hypothesis anchoring algorithm. Section 6 presents experimental results and we end with a conclusion and an outlook to future work in Section 7.

4. Anchoring

This section explains the anchoring framework as used in the PMHA algorithm. Because of the intuitive presentation in [9], our explanation of the anchoring framework in Section 4.1 is based on their framework using conceptual spaces. Section 4.2 explains the modifications we have made to the existing framework in order to allow multiple object tracking models. Fig. 1 gives a schematic overview of the complete algorithm and will be explained step-by-step in the following sections.

4.1. Anchoring framework

We start by defining the symbol system represented by the symbol block on the left in Fig. 1. First of all, the symbol system contains a set of individual symbols \( \Gamma = \{ \iota_1, \iota_2, \ldots \} \). Individual symbols represent instances, i.e., physical objects in the real world. In addition, the symbol system contains a set of predicate symbols \( \Pi \). Predicate symbols will be used to assign properties to individual objects, e.g., the individual symbol \( \iota_1 = \text{cup-12} \) might associate with the predicate symbol set \( \{ \text{cup}, \text{small1}, \text{blue} \} \).

In [9], a conceptual space is a metric space, where the dimensions are represented by measurable object attributes, e.g., HSV color, size, position. In this work, their conceptual space is referred to as predicate attribute space, since this better suits our terminology. Inside the rectangle on the bottom of Fig. 1.a two dimensional predicate attribute space with dimensions size and hue is visualized.

Predicate grounding relations \( g \) map predicate symbols to regions in the predicate attribute space, i.e., they map predicate symbols to measurable attribute values. Predicate grounding relations are defined to be time-invariant and shown in the bottom part of Fig. 1.

The time-invariant perceptual system is represented by the perception block on the right in Fig. 1. At each point in time, a set of object attributes is measured, e.g., a position and a color. At the discrete time step \( k \), the set of \( n_{\text{meas}} \) measurements that is generated by the perceptual system is represented by

\[
Z(k) = \{ z_i(k) \}_{i=1}^{n_{\text{meas}}}. \quad \text{The total set of measurements up to time } k \text{ is represented by } Z^k = \{ Z(j) \}_{j=1}^{k}. \text{ Each measurement can be represented by a point in the predicate attribute space, see Fig. 1. For reasons of compactness the time step } k \text{ will be written as a superscript.}
\]

In this work, we assume that the perceptual system contains routines that measure attribute values of objects with some uncertainty measure, e.g., position with a covariance matrix or a probability mass function (pmf) over different colors (Assumption 1). As a result, the point in the predicate attribute space representing a measurement extends to a region, represented by the blue circle in Fig. 1. Furthermore, each separate measurement \( z_i(k) \) in the set \( Z(k) \) is assumed to originate from at most one object (Assumption 2).

In the remainder of this paper, \( p(x) \) denotes a probability distribution over a random variable \( x \) and \( p(X) \) denotes the same distribution evaluated at the particular value \( x = X \).

4.2. Anchors

Real world objects are implemented by anchors, shown in the middle of Fig. 1. Anchors are maintained over time, however, their attributes will be updated based on new sensory information. Our anchors can be interpreted as a probabilistic version of the anchors in [8] including multiple model object trackers describing the object’s attributes:

\[
\alpha_k = (i, z^k_i, M^k_{a,i}), \tag{1}
\]

where \( i \) is an individual symbol, \( z^k_i \) a measurement performed at time \( k \) that will be linked to this symbol, \( i = 1, \ldots, n_{\text{meas}} \) is the measurement index and \( a = 1, \ldots, n_{\text{obj},b} \) is the anchor index. \( M \) contains the behavior models of the anchor:

\[
M_{a,m} = \{ p(M^k_{a,m,1}), M^k_{a,m}, \ldots \} \tag{2}
\]

\( M_{a,m} \) is a behavior model, \( n_{\text{mod},a} \) the number of behavior models in anchor \( \alpha_k \), \( m \) the model behavior index and \( p(M_{a,m}) \) the pmf over the behavior models evaluated at \( M_{a,m} \). The number of models \( n_{\text{mod},a} \) is assumed to be independent of \( k \) (Assumption III).

Within each behavior model \( M^k_{a,m} \), a pdf \( \gamma^k_{a,m} \) represents the attribute estimate according to the corresponding behavior model. In case of a linear KF, \( \gamma^k_{a,m} \) is the Gaussian distribution estimating the attribute of anchor \( \alpha_k \) at time \( k \). \( M^k_{a,m} \) is the complete KF including, e.g., the motion model. The final attribute estimate, referred to as the signature \( r^k_{a,m} \), is implemented as a convex sum:

\[
r^k_{a,m} = \sum_{m=1}^{n_{\text{mod},a}} p(M^k_{a,m}) \gamma^k_{a,m}, \tag{3}
\]

of the estimates of all behavior models in \( M^k_{a,m} \). In Fig. 1, the two red curves represent pdfs \( \gamma^k_{a,1} \) and \( \gamma^k_{a,2} \), estimating some attribute and \( p_1 \) and \( p_2 \) represent the relative weights of the corresponding behavior models \( p(M^k_{a,1}) \) and \( p(M^k_{a,2}) \). Section 6.1.1 shows how all relevant variables, relations and modules were selected during one of our experiments.

5. Probabilistic multiple hypothesis anchoring

In Section 4 we have explained how anchors are implemented. This section incorporates a data association algorithm thereby completing our probabilistic multiple hypothesis anchoring (PMHA) algorithm. The data association algorithm decides which anchors in the world model have to be updated using which measurements, i.e., data association allows correctly updating the anchor’s attributes.
Fig. 1. Schematic overview of the PMHA algorithm. At the top the hypothesis tree. Each dot represents a hypothesis and the red crosses represent branches that are pruned. Every hypothesis contains a list of anchors and has a probability of being correct. Each anchor on its turn contains an individual symbol, a set of measurements and a probabilistic signature that consists out of a mixture of probability density functions generated by a set of behavior models. The two behavior models are not shown, however, the pdfs generated by these behavior models with the corresponding weights $p_1$ and $p_2$ are represented by the red curves. The predicate attribute space represents predicate grounding relations that link attribute values and predicate symbols. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.1. Generating multiple hypotheses

In the PMHA algorithm, we include multiple hypothesis tracking-based data association. The main idea underlying the MHT is that each measurement can:

1. Represent an object not present in the world model.
2. Originate from a previously observed (existing) object.
3. Be a false detection (clutter).

Each possible set of associations from the $n_{\text{meas}}^k$ measurements received at time $k$ to anchors and/or clutter is represented by its own set of assignments $\theta^k$.

A hypothesis contains a list of anchors describing one possible state of the world and is denoted by $\Theta_h^k$, where $h = 1, \ldots, n_{\text{hyp}}$ is the hypothesis index. The parent of a hypothesis $\Theta_h^k$ is denoted $\Theta_{h-1}^{k-1}$, hence the set of assignments $\theta^k$ maps $\Theta_{h-1}^{k-1}$ to $\Theta_h^k$. By considering all possible associations, each parent node expands to multiple children each time new measurements arrive. This way a hypothesis tree is formed, which is schematically visualized in the upper part of Fig. 1.

5.2. Propagating, initiating and updating anchors

Each time a new measurement arrives all attributes of all anchors are propagated using the behavior models in $\mathcal{M}$, e.g., a KF propagation step. Furthermore, the probabilities $p(M_{a,m}^k)$ are propagated using transition probabilities among the different models in $\mathcal{M}$ as used in MM estimators [21]. To get a better understanding of the probabilities $p(M_{a,m}^k)$ after propagation, two examples are briefly discussed:

1. If a robot observes a mug standing on the dining table a typical behavior model in the model set used throughout this work represents the mug position by a constant velocity KF initialized with a zero velocity. In addition, a common sense based second behavior model is a uniform distribution over the table, since mugs are typically moved every now and then. During propagation $p(M_{a,1}^k)$ decreases whereas $p(M_{a,2}^k)$ increases.

2. For predicting people movements, e.g., predicting if John will be at work or at home, more advanced human mobility models are needed. Spatial and temporal based next place prediction as described in [24,25] can be included in the model set $\mathcal{M}$. However, incorporating such next place prediction is left for future work.

After the propagation step, the measurements can be associated with anchors or with clutter. In case of associating a measurement with a new object, a new anchor is initialized with a signature according to the measured attribute values. The behavior model
set $\mathcal{M}$ is object specific and loaded from a knowledge base. A default model set containing a constant velocity KF and a uniform distribution is used if an object is not represented in the knowledge base.

In case of associating a measurement with an existing anchor, the attributes are updated, e.g., a KF update step. The probability $p(M_{a,m})$ belonging to the model that is updated is set to one, the other probabilities are set to zero. During the propagation step, the behavior model probabilities belonging to the updated models will decrease whereas the other probabilities will typically increase.

In case of associating a measurement with a leaf of the track tree rather than the actual anchor, the behavior model probabilities are set to one. The predicted attribute value falls with a certain probability. The association between anchor and measurement is valid if the likelihood of the measurement given the prediction is calculated only if this likelihood is above some user defined threshold. If the anchor does not yet contain an estimate of the measured attribute, the association is always considered valid.

To avoid a brute force implementation that enumerates all legal associations, we have, like in [26], implemented a strategy that generates the $n_{\text{hyp}}$ most probable hypotheses only. This is an approximation since there is no guarantee that the actual solution is included in the set of $n_{\text{hyp}}$-best hypotheses. To further reduce the computational and storage overheads, we use track trees as in [26]. In the track trees all anchors are stored and the hypotheses contain pointers to the leaves of the track tree rather than the actual anchors.

Both these strategies imply that some of the hypotheses are not expanded. This is visually represented by the red crosses in Fig. 1.

### 5.3. Managing the tree growth

If all possible associations between anchors in the world model and measurements are considered, the size of the tree grows rapidly due to the combinatorial explosion. This section explains two strategies that are implemented to limit the tree growth.

Just like the predicate symbols and the measurements shown in Fig. 1, each pdf $\Gamma_i$ representing the attribute estimate of an anchor can be mapped to a region in the predicate attribute space in which the predicted attribute value falls with a certain probability. The likelihood of the measurement given the prediction is calculated only if this likelihood is above some user defined threshold, the association between anchor and measurement is valid. If the anchor does not yet contain an estimate of the measured attribute, the association is always considered valid.

### 5.4. Probabilities of hypotheses

The probabilities of the hypotheses being correct is calculated using Bayes’ law:

$$p(\Theta^k | Z^k) = \frac{p(Z(k) | \Theta^k, Z^{k-1}) p(\Theta^k | Z^{k-1})}{p(Z(k) | Z^{k-1})}, \quad (4)$$

where $p(\Theta^k | Z^k)$ is the posterior probability of the hypothesis with index $h$ given all measurements up to time step $k$, $p(Z(k) | \Theta^k, Z^{k-1})$ is the likelihood, i.e., the probability of the measurements performed at time step $k$ given the hypothesis, $p(\Theta^k | Z^{k-1})$ is the prior probability of the association set $\Theta^k$ given the corresponding parent hypothesis and all measurements up to time step $k - 1$ and $p(\Theta^k | Z^{k-1})$ is the posterior probability of the parent hypothesis. Finally, $p(Z(k) | Z^{k-1})$ is the normalization term. The following sections zoom in on the calculation of the likelihood and the prior.

#### 5.4.1. Likelihood

Under the assumption of independent measurements (Assumption IV), the likelihood factorizes to:

$$p(Z(k) | \Theta^k, Z^{k-1}) = \prod_{i=1}^{n_{\text{meas}}} p(z_i^k | \alpha^k_{i,a_i}, Z^{k-1}) \cdot \prod_{j=1}^{n_{\text{false}}} p(z_j^k | \Theta^k_{\text{false}}), \quad (5)$$

where $\alpha_i$ denotes the index of the anchor to which the measurement $z_i^k$ is associated in hypothesis $\Theta^k_i$.

In case of associating a measurement with a new object or with clutter, we assume the probability on the right hand side of Eq. (5) to be uniformly distributed over the observation region with volume $V$ (Assumption V), i.e., the probability of false detections and new objects is independent of the position of the measurement in $V$, and hence:

$$p(z_i^k | \alpha^k_{i,a_i}, Z^{k-1}) = V^{-n_{\text{const}}_{i,a_i} - n_{\text{false}}}, \quad (6)$$

where $n_{\text{const}}_{i,a_i}$ and $n_{\text{false}}$ are the number of new respectively false measurements according to the corresponding hypothesis. Assumption V is commonly made in data association literature [26] and can be relaxed if preferred.

In case of associating the measurement with an existing anchor, the probability on the right hand side of Eq. (5) can be calculated using $\Gamma^k_{\text{true}}$ since the signature summarizes all measurements associated with the anchor and the probability of observing $z_i^k$ only depends on the attributes of the anchor:

$$p(z_i^k | \alpha^k_{i,a_i}, Z^{k-1}) = p(z_i^k | \Gamma^k_{\text{true}}) = \prod_{m=1}^{n_{\text{mod}}(a_i)} p(M_{a,m}) \cdot p(z_i^k | \Gamma^k_{m}), \quad (7)$$

where the first term on the right hand side represents the probability of behavior model $M_{a,m}$ being correct and the second term represents the probability that the object underlying the anchor generates the measurement $z_i^k$ under the assumption that model $M_{a,m}$ indeed correctly describes the anchor’s attribute values. Again, $n_{\text{mod}}(a_i)$ is the number of behavior models in $\mathcal{M}$.

We have selected the probability $p(z_i^k | \Gamma^k_{m})$ to be dependent on the mismatch between the predicted and the measured attribute value. Evaluating this probability analytically is feasible for a restricted set of pdfs only, e.g., Gaussian pdfs. For that reason, we use pdfs from this class in the model set $\mathcal{M}$.

#### 5.4.2. Prior

The probability of the association set $\Theta^k$ given a parent hypothesis $\Theta^k_{p(h)}$ is, inspired by [26]:

$$p(\Theta^k | \Theta^k_{p(h)}, Z^{k-1}) = \frac{n_{\text{false}} n_{\text{true}}}{n_{\text{meas}}} \cdot \int p(v_{n_{\text{true}}}) \cdot p(n_{\text{false}}) \cdot p(n_{\text{true}}_{\text{false}}) \cdot p(n_{\text{true}}_{\text{true}}) \cdot \prod_{a=1}^{n_{\text{mod}}(h)} p(D^k_{a,m})^{\delta_a} (1 - p(D^k_{a,m}))^{1-\delta_a}, \quad (8)$$

where $p(n_{\text{true}})$ and $p(n_{\text{false}})$ are the prior pmfs of the number of new objects and the number of false detections and $p(D^k_{a,m})$ is the probability of detecting anchor $\alpha_a$ in hypothesis $\Theta^k$.

The indicator value is defined as:

$$\delta_a = \begin{cases} 1 & \text{if anchor } \alpha_a \text{ in } \Theta^k_{p(h)} \text{ is detected at time } k \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

Each anchor has a probability $p(V^k_{a,m})$ representing the probability of the object currently being visible. Under the assumption that objects that are not visible will not be detected (Assumption VI), $p(D^k_{a,m})$ can be written:

$$p(D^k_{a,m}) = p(D^k_{a,m} | V^k_{a,m}) = p(V^k_{a,m} | M^k_{a,m}) \cdot p(M^k_{a,m}) \cdot p(V^k_{a,m} | \neg M^k_{a,m}) \cdot p(\neg M^k_{a,m}), \quad (10)$$

Finally, the association set $\Theta^k$ is represented by:

$$p(\Theta^k | \Theta^k_{p(h)}, Z^{k-1}) = p(D^k_{a_m}) \cdot \prod_{m=1}^{n_{\text{mod}}(a_m)} p(M^k_{a,m}) \cdot p(V^k_{a,m}), \quad (11)$$
where \( p(D^k_{h,a} \mid V^k_{h,a}) \) is the probability of detecting an object given that it is visible and can be determined based on the reliability of the perceptual routines. The pmf \( p(M^k_{h,a,m}) \) is as introduced before and \( p(V^k_{h,a} \mid M^k_{h,a,m}) \) is the probability of the object being visible given that model \( M^k_{h,a,m} \) describes the correct object attribute values.

5.5. Generating output

The only step left is providing the robot with a final decision based on all possible world states in the tree. In this work we select the world state described by the most probable hypothesis.

6. Experiments

This section describes the experiments that were performed. The experiments in Section 6.1 focus on the anchoring part of our framework. Section 6.1 includes a detailed overview of all the components in Fig. 1 and shows when the PHMA algorithms fails due to the assumptions made. In Section 6.2, a robot is confronted with an ambiguous scenario in a typical household setting scenario. The focus is on the multiple hypothesis aspect of the PHMA algorithm. Section 6.3 presents an experiment in a highly dynamic environment and, therefore, focuses on data association and some of the probabilistic features of the proposed approach.

6.1. Dynamic objects on a table

The experiment presented in this section focuses on the anchoring aspect of the PHMA algorithm and has various purposes. It shows:

- Detailed examples of all components and models related to the PHMA algorithm.
- How the proposed approach fails if some of the assumptions are not met.
- How the semantically rich objects reduce the number of valid associations between measurements and world model objects.

6.1.1. Implementation example

In this experiment, a table is observed. Five different objects appear, all shown in Fig. 2. The predicate symbol set associated with each of these objects is shown in Table 1. Even though the two tea boxes shown in Fig. 2(a) and (b) are different in size, they are associated with the same set of predicate symbols. As a result, their appearance in the world model is identical. The same holds for the red objects shown in Fig. 2(d) and (e). The symbols system used in this experiment is inspired by the latest version of the original anchoring framework introduced in [11].

The predicate grounding relations are shown in Table 2. In this table, size is the length of the diagonal of the detected 2D blob in meters and ratio is the height of the blob divided by its width.

During this experiment, the perceptual system only contains a 2D blob detector that determines the color, size and shape of the blobs according to the predicate grounding relations. Depth information is used to estimate a 3D position and to determine the length of the diagonal. An example measurement could look like this:

\[
\begin{align*}
\text{size} &= \text{small} (p = 0.8) \\
\text{shape} &= \text{rectangle} (p = 0.8) \\
\text{color} &= \text{red} (p = 0.8) \\
\text{position} &= (1.02, 0.23, 0.87) \times (\Sigma = \text{diag}(0.0008, 0.0008, 0.0008)).
\end{align*}
\]

For now, the probabilities of the measurements are fixed to 0.8 and the position covariance matrix \( \Sigma \) as shown in this example. The probability of a measurement being clutter is set to 0.01. The probability of a new object is varied but always lower than the probability of clutter. In our experiments, these settings result in associating measurements with objects only if the set of discrete attribute values are identical. New objects appear only after multiple observations since the hypothesis associating a single measurement originating from a new object with clutter instead is preferred. This proved reasonable during all of the presented experiments. The models are defined in a separate knowledge base, which makes them easy to adapt. Learning the various probabilities from data or deriving the position covariance based on sensor characteristics could lead to better choices, however, this is left for future work.

The behavior model set \( \mathcal{M} \) in this experiment contains two models for all objects. A constant velocity \( KF \) initialized with a zero velocity at the location of the first measurement associated with the object and a uniform distribution over the table area. During propagation, the position uncertainty increases rapidly due to expected random object displacements. An occlusion of about five seconds is sufficient to switch to the uniform distribution and as a result, measurements anywhere on the table cause a valid association with any object on the table with the same predicate symbol set. The predicate symbol set associated with an object is modeled to be time invariant in this experiment.

6.1.2. Results and the effect of key assumptions

In the first part of the experiment, three objects are present. One of the objects is moved around. Fig. 3 shows the three objects with typical detections generated by the blob detector and the camera input with world model overlay. In the world model overlay, the shape, size, color, position and unique ID are displayed. All objects have a unique set of attribute values which simplifies data association in this part of the experiment.

In the second part of the experiment, the full scene is occluded for about four seconds. During the occlusion, object positions are propagated using the model set \( \mathcal{M} \). As a result, the probability \( p(M^k_{h,a}) \) of the first behavior model decreases and the probability of the uniform distribution \( p(M^k_{h,a,m}) \) increases. This is a design decision that models that the objects can be moved to any location on the table.

Fig. 4(a) shows the world model immediately after the occlusion. From this figure, it can be observed that the red squared tea box present in Fig. 3(b) is removed during the occlusions and the red cup is added. Fig. 4(b) shows that the perceptual system generates false negatives regularly. In Fig. 4(c), the conclusion of the world model is shown. Both the red cup and the red
tea box are associated with the same set of predicate symbols. As a result, associating measurements originating from the cup with the world model object that represented the red tea box is valid. The measurements of the red cup are incorrectly associated with the object with ID 2. The main conclusion drawn from this association error is that summarizing attributes into a set of discrete values might cause information loss and, thereby, complicates data association.

In the last part of this experiment, the object shown in Fig. 2(b) is added to the scene as shown in Fig. 5(a). This does not cause any problems in the world model. Then, this new object is moved towards the other rectangular green tea box. As a result, the 2D blob detector switches between the output shown in Fig. 5(b) where the detections are correct and Fig. 5(c), where one measurement of a large green rectangle is measured instead. The large green rectangle does not associate with the green medium sized rectangles in the world model and as a result, the world model adds a large green rectangle as shown in Fig. 5(d). The reason of this failure is Assumption II, i.e., a measurement is assumed to originate from at most one object. The possibility of two objects generating one measurement together is not considered.

The independence of measurements assumed in Assumption IV makes the probability of an object being present increase rapidly after receiving a number of false positives with the same attribute values. As a result, the hypothesis describing these measurements as being clutter gets a low probability compared to the hypothesis shown in Fig. 5(d). Dropping Assumption II is possible, however, this implies a higher number of possible explanations per measurement set. Therefore, it adds computational complexity and requires additional models modeling this effect.

6.2. Ambiguities with cups

Robots in domestic environments will inevitably be confronted with ambiguous scenarios. By maintaining multiple hypotheses and calculating probabilities for each of them, the PMHA algorithm can deal with those scenarios. The purpose of this experiment is to show:

- The benefit of maintaining multiple hypotheses.
- The flexibility offered by the probabilistic framework and the corresponding probabilistic models.
- The advantages of an anchoring framework over a framework without anchoring.

6.2.1. Scenario and object recognition

In this experiment the robot will face a table that contains three objects, a tea can, a tea box and a cup, see Fig. 6(a). Each of these three objects is present in the world model right from the start of the experiment. Then the full scene will be occluded, see Fig. 6(b). After a while, the occlusion becomes partial, see Fig. 6(c). During this partial occlusion one cup is observed, however, not at the location of the cup observed prior to the full occlusion. This ambiguous scenario will be resolved after the full scene is observable again, but it is up to the world modeling algorithm to make an association decision already during the partial occlusion.

During this experiment, the measured attributes are position and a class label, either cup, tea-can or tea-box. As new measurements arrive, the probability of the anchor’s attribute value being tea-can increases as new tea-can measurements arrive around the predicted location. As a result, measurements with different labels will not be associated with the anchor with class label attribute tea-can anymore, as explained in Section 5.3. The same holds for the other objects. The predicate grounding relations relevant for this experiment link viewpoint feature histograms (VFHs) describing the 3D shape of a cluster of points in a point cloud to the class labels cup, tea-can and tea-box. See [27] for more details about VFHs.

6.2.2. Results

We have performed the experiment described above twice. In the first experiment, we have set the probabilistic models such
that we have a high probability of new cups entering the scene and we track the position attribute of the anchors using a KF with a constant position motion model. As a result, the most probable hypothesis will be the one explaining the cup during the partial occlusion as being a new cup. Once the complete scene is observed again, this appears to be right, as can be seen in Fig. 7, where the point cloud observed by the robot is shown together with text labels indicating the anchor’s class label in the world model and unique IDs that allow distinguishing various instances of the same class.

In the second trial, we have kept everything the same except for two things. We have (i) lowered the probability of new cups entering the scene and (ii) added a second model to the behavior model set $M$. This second model represents an increasing probability of cups being moved to elsewhere on the table over time. Changing the class dependent model set and probabilities can be done in a separate configuration file.

Fig. 8 shows the results with these adapted probabilistic models. During the partial occlusion shown in Fig. 8(b), the most probable hypothesis is the one that assumes that the cup was moved during the full occlusion. The hypothesis representing the output that was given during the previous experiment is still present, however, because of the probabilistic models we have selected, that hypothesis has a lower probability. After the occlusion being resolved it becomes apparent that the original cup was not moved but a new cup was added during the occlusion. The most probable hypothesis changes hence the PMHA algorithm corrects its previous conclusion based on new evidence as shown in Fig. 8(c). During the partial occlusion it is impossible to know which of the two alternatives represents the true solution. Besides the advantage of maintaining multiple hypotheses, this experiment also demonstrates the ability of incorporating various kinds of prior knowledge in the probabilistic object tracking models.

6.3. Ant tracking

The Georgia Institute of Technology College of Computing recorded sixteen videos of ants in the process of nest migration. “The artificial nest located in the field of view consists of a cavity constructed out balsa wood and a top covering pane of glass. Ants may enter the nest in the entrance at the top of the image, but they
Fig. 8. Output of the PMHA algorithm with a relatively high probability of new cups entering the scene.

Fig. 9. Illustrative frame of video 1 with the detected blobs. The red lines indicate the rectangular bounding boxes around the blobs, the crosses the ant position approximated by the blob centers. An occlusion in (a) results in an inaccurate position estimate. The effect of these detections on the world model generated by the PMHA algorithm is shown in (b).

6.3.1. Ant detection and object behavior models

In this experiment we associate the individual symbols in the anchors in the world model with the predicate symbol brown. The predicate grounding relation for the predicate symbol brown defines the color in the YUV color space by $Y \in [39, 101]$, $U \in [116, 125]$, and $V \in [128, 136]$ as suggested in [28].

The perception module detects brown blobs in the downscaled image, again as suggested in [28], and the positions of these blobs. The ant location is assumed to be the centre point of the smallest possible rectangular bounding box enclosing the blob. Fig. 9 shows two frames from video 1. The red lines represent the bounding boxes around the blobs that were detected, the red crosses represent the positions of the blobs. Fig. 9(a) shows that our basic detection algorithm generates inaccurate position estimates in the presence of occlusions. The effect of such detections on the world model is shown in Fig. 9(b), where it can be seen that the light blue cross is a little bit off due to the inaccurate approximation of the position. Better results can be obtained by adding predicate symbols like shape or size, however, that would no longer allow a fair comparison with other works using the same data set [28, 5].

The behavior models set includes a linear KF with a constant velocity motion model only.

6.3.2. Results

The PMHA algorithm was successful in twelve out of the sixteen videos. The data association failed for videos 3, 8, 12 and 14, i.e., ants identities got mixed, which is understandable given the fact that the constant velocity motion models were very poor estimates for the actual ant movements in these videos. In two out of the twelve successes, the hypothesis tree corrected previous conclusions based on new evidence. The performance is comparable to state of the art MTT algorithms such as, [28, 5], that also failed in four out of the sixteen videos. However, they excluded the anchoring part that we think is inevitable in a domestic robotics setting. A more detailed comparison of the results is shown in Table 3, copied from [5] and extended with both the results from [28] and our results. Anchoring algorithms without probabilistic data association will fail in dynamic scenarios including ambiguities as presented in this section. Anchoring with GNN data association will perform worse. The average computation time the PMHA algorithm needed per image was 29 ms on a HP EliteBook 8530w with Intel 2.8 GHz duo core for a non-optimized implementation, which allows real-time execution.

Fig. 10 shows an example of the estimated ant trajectories for videos 1 and 6. All measurement to ant associations are correct. Complicating factors in these particular videos are an occlusion causing 13 subsequent missed detections of the same ant in video 1 and various occlusions among various ants in video 6. The somewhat wobbly trajectories are the result of the inaccurate position estimation during occlusions, as visualized in Fig. 9(a).

Each hypothesis has a probability and each hypothesis represents a possible world state with a certain number of ants in the observation region. We have calculated the expected value of the number of ants based on all hypotheses in the tree and their probabilities. For videos 1 and 6, the expected value for the number of ants according to the hypothesis tree is plotted over time in Fig. 11. The red dots indicate the number of detections at each of the frames. Even though occlusions and the poor detection algorithm cause many false negatives, the PMHA algorithms is able to correctly estimate the varying number of ants over time. This mainly is the result of the behavior models used for target tracking.
Fig. 10. Estimated trajectories of the ants in video 1 in Figure (a) and video 6 in Figure (b).

Fig. 11. Expected value of the number of ants according to the PMHA algorithm (black line) and the number of detections (red dots) for video 1 in (a) and video 6 in (a). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

| Source: Partly reproduced from Ref. [5]. |
| NOA | NOA entering | NOA leaving | Failure |
| Video 1 | 8 | 1 | 0 | No | No | No |
| Video 2 | 10 | 3 | 1 | No | No | No |
| Video 3 | 12 | 1 | 0 | No | No | Yes |
| Video 4 | 10 | 1 | 0 | No | No | No |
| Video 5 | 12 | 3 | 0 | No | Yes | No |
| Video 6 | 12 | 3 | 1 | No | No | No |
| Video 7 | 11 | 2 | 0 | No | No | No |
| Video 8 | 10 | 0 | 1 | Yes | Yes | Yes |
| Video 9 | 10 | 2 | 0 | No | No | No |
| Video 10 | 11 | 0 | 0 | No | No | No |
| Video 11 | 8 | 1 | 0 | No | No | No |
| Video 12 | 10 | 4 | 1 | Yes | Yes | Yes |
| Video 13 | 11 | 0 | 0 | No | No | No |
| Video 14 | 6 | 1 | 1 | Yes | Yes | Yes |
| Video 15 | 6 | 0 | 0 | No | No | No |
| Video 16 | 10 | 1 | 1 | Yes | No | No |

The behavior model set \( \mathcal{M} \) only includes low level motion models for tracking the position attribute in this example, however, an extension with other kinds of behavior models and different attributes is straightforward, as shown in the previous experiments.

7. Conclusions and future work

In this work, an approach called probabilistic multiple hypothesis anchoring (PMHA) is introduced and used for world modeling in a robotics setting. PMHA incorporates the strengths of probabilistic data association and multiple model tracking into an anchoring algorithm, resulting in a probabilistic anchoring approach. The approach on the one hand allows semantically rich objects with many different attributes and on the other hand allows correcting previous data association decisions based on new evidence which is particularly useful in ambiguous situations. The traditional multiple model target tracking was generalized such that various types of prior knowledge can be incorporated and the real-time applicability was demonstrated during a series of experiments.

The first experiment gives a detailed description of the various components and models in the proposed PMHA framework. It in addition demonstrated how semantically rich object descriptions are exploited while solving the data association problem and the most important assumptions affect the performance. The second set of experiments showed the (i) ability to correct previous data association decisions based on new evidence, (ii) advantages of tracking multiple attributes as it is done in anchoring algorithms and (iii) the effect of incorporating prior knowledge by means of multiple behavior models for object attribute tracking. We think the combination of these features is very important since robots will be confronted with ambiguous situations regularly. The third set of experiments showed the ability to perform correct association with a varying number of visually identical objects, non-linear trajectories, missed detections and full occlusions which is nearly impossible with anchoring algorithms that do not include probabilistic data association. The probabilistic information available in the hypothesis tree was used to show both the most probable state of the world and the expected number of ants in the scene.

Future work will be to further investigate the probabilistic models that are used. A valuable extension would be to autonomously
learn these possibly temporal or spatial dependent models from data, e.g., starting from [24,25]. Combining the PMHA algorithm with a reasoning module paves the way for a probabilistic reasoning module that reasons about various states of the world. The data present in our PMHA algorithm allows answering queries regarding the world state in a probabilistic manner, e.g., "Where is John?" can be answered with "at the office (p = 0.68)".

Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Program FP7/2007–2013 under grant agreement no. 248942 RoboEarth. We would like to thank the anonymous reviewers for their useful suggestions and their valuable feedback.

References


