Hypotheses Disambiguation in Retrospective

Ohad Shelly and Vadim Indelman, *Member, IEEE*

***Abstract*—Robust perception is a key required capability in robotics and AI when dealing with scenarios and environments that exhibit some level of ambiguity and perceptual aliasing. In this work we consider such a setting and contribute a framework that enables to update probabilities of externally-defined data association hypotheses from some past time with new information that has been accumulated until current time. In particular, we show appropriately updating probabilities of past hypotheses within this smoothing perspective potentially enables to disambiguate these hypotheses even when there is no full disambiguation of the mixture distribution at the current time. Further, we develop an incremental algorithm that re-uses hypotheses' weight calculations from previous steps, thereby reducing computational complexity. In addition we show how our approach can be used to enhance current-time hypotheses pruning, by discarding corresponding branches in the hypotheses tree. We demonstrate our approach in simulation, considering an extremely aliased environment setting.**

***Index Terms*—Localization, mapping, SLAM.**

I. Introduction

Autonomous navigation in uncertain or unknown environments is essential in numerous applications in robotics, such as search and rescue, autonomous cars, indoor navigation, and surveillance. Once the robot operations take place in an unknown or uncertain environments, the navigation process also involves environment mapping. The corresponding problem, known as simultaneous localization and mapping (SLAM), has been extensively investigated [1] in the last two decades by the robotics and computer vision communities. Computationally efficient online solvers that exploit the underlying inherent sparsity of the problem and re-use of calculations are readily available [2]–[4].

Traditional SLAM approaches include two parts, commonly known as the "front-end" and the "back-end". The latter maintains and updates a belief over robot past and current states (e.g. poses) and mapped environment given the available data at each time instant. This data can include any prior information, if exists, performed actions and captured sensor observations with the corresponding data association (DA). The latter is determined by the front-end process, and can be considered as associating observed scenes (e.g. in terms of landmarks) from current and previous time instances. A correct association between an observed landmark and a received measurement is crucial for accurate inference.

A key common assumption is that the data association has been correctly determined by the front-end. Such an assumption, however, is less valid in presence of perceptual aliasing and ambiguity. An incorrect data association can lead to catastrophic results in inference/SLAM, e.g. the robot might deduce it is located in an incorrect similar-looking corridor, while assuming it is perfectly solved within planning can lead to sub-optimal actions, that could lead to collision and unsafe behavior, in general.

Relaxing the data association assumption would lead to robust perception approaches that are much required while operating in the real world [1], which typically exhibit some level of perceptual aliasing. Yet, this involves reasoning about DA as part of inference, and results in a set of hypotheses, where each one is built by a possible landmark association to the given measurement in hand. Such a formulation corresponds to a multi-modal belief, that can be represented, e.g. by a Gaussian mixture model (GMM) [5], [6].

While robust inference approaches have been actively investigated in the last few years, till now existing approaches have dealt with relaxing the DA assumption while examining the state distribution at the *current* time instant. Moreover, except of [7], typically calculations are done from scratch for each time step, without calculation re-use.

In contrast, in this work we propose the notion of *hypotheses disambiguation in retrospective*, i.e. after more information has been collected. Our approach enables to re-evaluate the probability of externally-defined, key strategic hypotheses from a past time, given the information obtained up to the current time, while accounting for the data association hypotheses developed since that past time. See illustrations in Fig. 1. For example, these hypotheses could refer to an observation of an important but ambiguous event (e.g. did we observe scene/object A or B in the past?). We propose to utilize data that has been obtained since that time to update the posterior probabilities of these key past hypotheses. We envision such a capability and the general concept to be of interest in various contexts in robotics and beyond.

A. Related Work

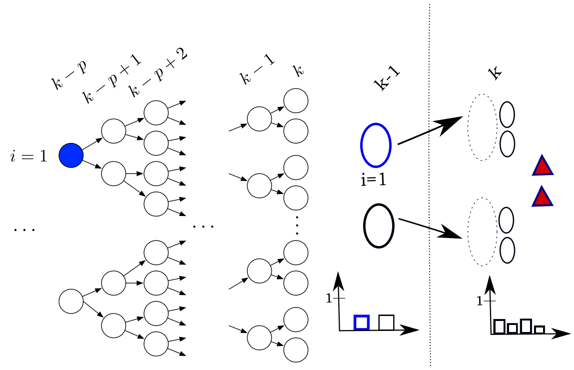
In the past years the research community has been actively investigating robust inference approaches to be resilient to false data association overlooked by front-end algorithms, i.e. by relaxing the assumption that the DA provided by front-end algorithms is outlier-free. An early work on DA is joint probability data association (JPDA) by Fortman *et al.* [8] which considers all possible DA options in the context of multi-target tracking. Sunderhauf and Protzel [9] introduced the so called

Manuscript received September 8, 2021; accepted December 30, 2021. Date of publication January 14, 2022; date of current version January 26, 2022. This letter was recommended for publication by Associate Editor G. Grisetti and Editor S. Behnke upon the evaluation of the reviewers' comments. This work was supported by the U.S. NSF/US-Israel BSF. *(Corresponding author: Ohad Shelly.)*

Ohad Shelly is with the Technion Autonomous Systems Program (TASP), Technion-Israel Institute of Technology, Haifa 32000, Israel (e-mail: ohshely2010@gmail.com).

Vadim Indelman is with the Department of Aerospace Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel (e-mail: vadim.indelman@technion.ac.il).

Digital Object Identifier 10.1109/LRA.2022.3143298



(a) General problem illustration (b) Single step update setting

Fig. 1. (a) In this work we aim to calculate the probability of externally-defined hypothesis from some past time $k-p$ given the information that has been obtained until current time k. In other words, we would like to calculate $w^i_{k-p|k} = \mathbb{P}(\gamma_{k-p} = i \mid H_k)$. The diagram illustrates, for simplicity, a branching factor of two, i.e. each hypothesis branches at the next time into two child hypotheses (e.g. due to obtained measurement with an ambiguous data association). For instance, we might be interested in calculating the weight $w^{i=1}_{k-p|k}$, which corresponds to probability of the hypothesis indicated by the **blue** node, given data until time k. (b) A toy example illustration of the general concept from (a), where $p = 1$, the prior belief is a Gaussian Mixture with two components, and upon performing an action the robot makes an observation z_k of one of two identical triangular landmarks. Due to ambiguous data association, the number of components in the posterior belief at time k is four. Posterior and propagated beliefs are denoted by solid and dashed lines, respectively. In this single step update scenario we wish to re-evaluate at time k the weight of hypothesis i from time $k-1$, i.e. $w^i_{k-1|k}$, given the observation z_k.

switchable constraints to detect faulty loop closures that lead to erroneous data association in back-end optimization. Olson and Agarwal [10] proposed a robust approach that uses max-mixture models. Carlone *et al.* [11] addressed the problem from a different perspective, looking for a maximal coherent set among the given loop closure candidates. Indelman *et al.* [12] proposed a multi-robot framework for SLAM with ambiguous data association. Wong *et al.* [13] presented a Dirichlet Process Mixture Model (DPMM) for data association in partially observed environments. More recently, optimization approaches robust to outliers have been investigated in works such as [14], [15]. Finally, [16] and [7] develop incremental optimization approaches considering ambiguous data association.

A recent work by Pathak *et al.* [5] targets perceptual aliasing by explicitly reasoning about and probabilistically maintaining ambiguous DA hypotheses, in both inference and belief space planning. In contrast to many of the works mentioned above, it explicitly calculates the probability of different hypotheses, i.e. weights of GMM components, rather than assuming these to be identical. Tchuiev *et al.* [6] extend the passive inference formulation from [5] by utilizing semantic information and viewpoint-dependent classifier models, as well as weight pruning to reduce the number of DA hypotheses.

B. Contributions

While the above-mentioned approaches address inference considering the (GMM) belief from the current time, we investigate a complimentary aspect, namely, utilizing current information to re-evaluate the probability of past, externally-specified, DA hypotheses. Specifically, building upon [5], [6] we develop a smoothing approach for updating the weights of past GMM beliefs, while properly accounting for the ambiguous data association hypotheses that have been acquired since then.

Specifically, our main contributions in this letter are as follows: (a) we introduce the problem of hypotheses disambiguation in retrospective, which, to the best of our knowledge has not appeared thus far in literature; (b) we develop a probabilistic approach to update the probability of selected past hypotheses considering a smoothing formulation, while properly accounting for the ambiguous data association hypotheses that have been acquired since that time; (c) we derive a scheme for calculation re-use within this approach to reduce computational time; (d) we enhance hypotheses pruning also at *current* time, by leveraging the proposed concept of past hypotheses re-evaluation in retrospective and drawing a connection between hypotheses at current time and the corresponding ancestor hypotheses; and (e) we evaluate our approach in simulation considering an extremely aliased environment comprising identical landmarks. This letter is accompanied with supplementary material [17].

II. Notations and Problem Formulation

A. Notations

Let x_k represent the robot current pose at time k, and denote by $X_k = \{x_0, x_1, \ldots, x_k\}$ all robot poses until that time. We define u_k and z_k to be, respectively, the robot's action and captured observation at time k. Further, we represent the environment by landmarks $L = \{l_i\}^{|L|}_{i=1}$, and assume they are static and known. We note that in case landmarks are uncertain, as in a typical SLAM setting, the observed landmarks up to time k would become part of the state X_k, and the formulation of our approach can be straightforwardly adjusted to such a setting. We should note, however, that such an extension will necessitate coping with the inherent curse of dimensionality, which is outside the scope of this letter. Furthermore, the concept presented in this work is applicable also to other environment representations, such as grid-based localization, as long as one can formulate the corresponding data association hypotheses (see e.g. [5]).

Further, denote data association (DA) at time k by a discrete latent variable β_k, i.e. measurement z_k is associated to landmark (or object/scene) $l_{\beta_k=p}$, where $p \in [1, |L|]$. It is important to note that in the general case of obtaining a number of measurements in a single step, β_k would have been addressed to as a vector built from the landmarks associated to each measurement, as done in [6].

We use motion and observation models

$$x_k = g(x_{k-1}, u_{k-1}) + w \ \ , \ \ z_k = h(x_k, l_{\beta_k}) + v, \qquad (1)$$

where $w \sim \mathcal{N}(\mu_w, \Sigma_w)$ and $v \sim \mathcal{N}(\mu_v, \Sigma_v)$. The process and measurement covariance matrices, Σ_w and Σ_v, as well as the functions $g(.)$ and $h(.)$ are assumed to be known.

Let history H_k represent all the robot's actions and received measurements till time k along with the set of known landmarks L, $H_k = \{z_{1:k}, u_{0:k-1}, L\}$. Similarly, denote by H^-_k history without the received measurement at time k, $H^-_k = \{z_{1:k-1}, u_{0:k-1}, L\}$. The probability density function (pdf), the *belief*, at time k over X_k is then given by,

$$b[X_k] = \mathbb{P}(X_k \mid H_k). \qquad (2)$$

Since we consider ambiguous scenarios, one cannot assume data association to be given and perfect. Similar to [5], [6], the number of hypotheses at time k, without yet considering pruning or merging, is given by all possible realizations of the sequence $\beta_{1:k} = \{\beta_1, \ldots, \beta_k\}$. For convenience we denote $\gamma_k \triangleq \beta_{1:k}$, and consider at time k to have M_k hypotheses, i.e. $\gamma_k \in [1, M_k]$. Thus, the ith hypothesis, i.e. $\gamma_k = i$, corresponds to a specific sequence of $\beta_{1:k}$.

Hence, by marginalization of (2) over γ_k, and chain rule,

$$b[X_k] = \sum_{i=1}^{M_k} \underbrace{\mathbb{P}(\gamma_k = i|H_k)}_{} w_k^i \cdot \underbrace{\mathbb{P}(X_k \mid \gamma_k = i, H_k)}_{} b^i[X_k], \tag{3}$$

where $b^i[X_k]$ and w_k^i represent, respectively, the conditional belief and the weight of the ith hypothesis at time k, and $\sum_{i=1}^{M_k} w_k^i = 1$. Finally, we denote the propagated belief as the belief conditioned on H_k^- instead of H_k, i.e. without considering the measurement at the current time, z_k. Similarly, a propagated belief for the ith hypothesis is defined as $b^{i-}[X_k] = \mathbb{P}(X_k \mid \gamma_k = i, H_k^-)$.

B. Problem Formulation

In this work we wish to re-evaluate, in retrospective, the probability of externally-specified hypothesis (or hypotheses) from some past time, i.e. given new information acquired since then. Specifically, we wish to re-evaluate a hypothesis weight for a given $\gamma_{k-p} = i$. Thus, our goal is to calculate

$$w_{k-p|k}^i \triangleq \mathbb{P}(\gamma_{k-p} = i \mid H_k), \tag{4}$$

where $1 \leq p < k$. In other words, in this work we investigate a smoothing perspective considering discrete random variables (data association hypotheses), which are, however, coupled with continuous random variables (e.g. robot poses).

Another variant of this problem is to re-evaluate the probability of some past association, β_{k-p} instead of a sequence of associations $\gamma_{k-p} \equiv \beta_{1:k-p}$. We consider this setting in Section III-D.

We believe both problem variants can be of interest in different contexts: For example, considering specific realizations of γ_{k-p} may be useful in terms of localization, as each such realization corresponds to a posterior over X_k, see (3); This is in contrast to considering specific data association realizations from past time $k-p$, i.e. β_{k-p}, which can be of interest on its own.

To shorten notations in the sequel, we denote $m \triangleq k-p$ and re-write (4) as $w_{m|k}^i \triangleq \mathbb{P}(\gamma_m = i \mid H_k)$.

III. Approach

A. Derivation of a General Formulation for $w_{m|k}^i$

In this section we develop a general formulation for calculating (4). First, we perform Bayes rule considering the most recent measurement, z_k:

$$\mathbb{P}(\gamma_m = i \mid H_k) = \frac{\mathbb{P}(z_k \mid \gamma_m = i, H_k^-)}{\mathbb{P}(z_k \mid H_k^-)} \cdot \underbrace{\mathbb{P}(\gamma_m = i \mid H_{k-1})}_{} w_{m|k-1}^i,$$

where we use, here and in the sequel, the fact $\mathbb{P}(\gamma_m = i \mid H_{k-1}, u_{k-1}) \equiv \mathbb{P}(\gamma_m = i \mid H_{k-1})$, i.e. the weights of a GMM are not impacted by the motion model.

Considering $w_{m|k-1}^i$, we now repeat the above process and perform Bayes rule once again, which yields,

$$\mathbb{P}(\gamma_m = i \mid H_k) = \frac{\mathbb{P}(z_k \mid \gamma_m = i, H_k^-)}{\mathbb{P}(z_k \mid H_k^-)} \cdot \frac{\mathbb{P}(z_{k-1} \mid \gamma_m = i, H_{k-1}^-)}{\mathbb{P}(z_{k-1} \mid H_{k-1}^-)} \cdot \underbrace{\mathbb{P}(\gamma_m = i \mid H_{k-2})}_{} w_{m|k-2}^i \tag{5}$$

It is not difficult to see that performing Bayes rule sequentially in a similar fashion yields the following formulation:

$$\boxed{w_{m|k}^i = \left[\underbrace{\frac{\prod_{j=0}^{p-1} \mathbb{P}(z_{k-j} \mid \gamma_m = i, H_{k-j}^-)}{\prod_{j=0}^{p-1} \mathbb{P}(z_{k-j} \mid H_{k-j}^-)}}_{} \Psi_{m|k} \right] \cdot w_{m|m}^i} \tag{6}$$

Here, $w_{m|m}^i$ is the weight of the i-th GMM component at time m, while $\Psi_{m|k}$ is the update factor that is based on the data obtained in the period $[m+1, k]$. Therefore, we need now to calculate this term in order to get $w_{m|k}^i$.

Since the denominator in (6) is not conditioned on $\gamma_m = i$, its explicit calculation can be avoided. Instead, we first calculate the numerator,

$$\tilde{w}_{m|k}^i \triangleq \prod_{j=0}^{p-1} \mathbb{P}(z_{k-j} \mid \gamma_m = i, H_{k-j}^-) \cdot w_{m|m}^i, \tag{7}$$

for all $i \in [1, M_m]$, i.e. all hypotheses from time m, and then normalize as

$$w_{m|k}^i = \tilde{w}_{m|k}^i / \sum_{q=1}^{M_m} \tilde{w}_{m|k}^q. \tag{8}$$

Calculating (6) requires first computing the terms $\mathbb{P}(z_{k-j} \mid \gamma_m = i, H_{k-j}^-)$ for all $j \in [0, p-1]$. In the next section we develop an approach to do so. Yet, naïvely, one would calculate each of the above terms from scratch. In contrast, in Section III-C, we derive an incremental version, which re-uses calculations from previous steps.

B. Single Step Update, $p = 1$

We start with the simplest case of a single step update, i.e. $p = 1$ and $m = k-1$, as illustrated in fig 1(b). In this case, the un-normalized weight from (7) is given by

$$\tilde{w}_{m|k}^i = \mathbb{P}(z_k|\gamma_m = i, H_k^-) \cdot w_{m|m}^i. \tag{9}$$

To calculate $\mathbb{P}(z_k|\gamma_m = i, H_k^-)$ we marginalize over robot pose at time k and all the possible landmark associations β_k for the measurement z_k. Applying chain rule yields

$$\mathbb{P}(z_k|\gamma_m = i, H_k^-) = \sum_{g=1}^{|L|} \int_{x_k} \mathbb{P}(z_k|\beta_k = g, x_k, \gamma_m = i, H_k^-) \cdot \mathbb{P}(\beta_k = g \mid x_k, l_g) \cdot \underbrace{\mathbb{P}(x_k \mid \gamma_m = i, H_k^-)}_{b_m^{i-}[x_k]} dx_k, \tag{10}$$

where $\mathbb{P}(\beta_k = g|x_k, \gamma_m = i, H_k^-) \equiv \mathbb{P}(\beta_k = g \mid x_k, l_g)$ indicates the probability of observing the landmark l_g from robot pose x_k. Here, and throughout the letter, we use $\square^i_m$ to indicate conditioning on $\gamma_m = i$.

We reiterate that, while in this letter we assume landmarks are known, in case landmarks are uncertain and part of the belief, we would need to also marginalize over them in (10). Furthermore, for simplicity in this work we model $\mathbb{P}(\beta_k = g \mid x_k, l_g)$ as a uniform distribution with some finite support Ω_{l_g} (i.e. only for certain viewpoints, a given landmark is within sensor's field of view), and re-write (10) as

$$\mathbb{P}(z_k|\gamma_m = i, H^-_k) = \sum_{g=1}^{|L|} \frac{1}{|\Omega_{l_g}|} \int_{x_k \in \Omega_{l_g}} \mathbb{P}(z_k \mid l_g, x_k) b^{i-}_m[x_k] dx_k. \quad (11)$$

Since an analytical solution is not feasible, we resort to a sampling based approach and approximate (11) considering a set of $\{x^{n,i}_k\}^S_{n=1}$ sampled values from $b^{i-}[x_k]$, with S denoting the number of samples:

$$\mathbb{P}(z_k|\gamma_m = i, H^-_k) \approx \frac{1}{S} \sum_{n=1}^{S} \sum_{g=1}^{|L|} \frac{1}{|\Omega_{l_g}|} \mathbb{P}(z_k|l_g, x^{n,i}_k) \mathbb{1}_{\Omega_{l_g}}(x^{n,i}_k), \quad (12)$$

where $\mathbb{1}_{\Omega_{l_g}}(x)$ is an indicator function, indicating if landmark l_g is within sensor's field of view from pose x.

Finally we calculate $w^i_{k-1|k}$ via normalization as in (8).

C. Multiple Steps Incremental Calculation of $w^i_{m|m+p}$

In the previous section we addressed the calculation of a single step update considering $p = 1$. In this section we consider the general case of $1 \leq p < k$, as illustrated in Fig. 1(a), and develop a formulation to calculate the terms $\mathbb{P}(z_{k-j} \mid \gamma_m = i, H^-_{k-j})$ from (6) for all $j \in [0, p-1]$. While a naïve approach would calculate each of these terms from scratch, we develop an incremental version that allows to re-use calculations between different values of j.

We start by considering $j = p-1$ and $j = p-2$, and then discuss calculations for a general $j \in [0, p-1]$.

1) $j = p-1$: The corresponding term in (6) for $j = p-1$ is $\mathbb{P}(z_{m+1}|\gamma_m = i, H^-_{m+1})$. One can observe this is identical to the single step update case considered in Section III-B, yet here we consider a single step from $m = k-p$. Therefore, following a similar process we generate a set of samples $\{x^{n,i}_{m+1}\}^S_{n=1}$ from the propagated belief $b^{i-}_m[x_{m+1}]$,

$$b^{i-}_m[x_{m+1}] \triangleq \mathbb{P}(x_{m+1} \mid H^-_{m+1}, \gamma_m = i), \quad (13)$$

and approximate $\mathbb{P}(z_{m+1}|\gamma_m = i, H^-_{m+1})$ as

$$\frac{1}{S} \sum_{n=1}^{S} \sum_{g=1}^{|L|} \frac{1}{|\Omega_{l_g}|} \mathbb{P}(z_{m+1}|l_g, x^{n,i-}_{m+1}) \mathbb{1}_{\Omega_{l_g}}(x^{n,i}_{m+1}), \quad (14)$$

which corresponds to (12). Note that for this first step, i.e. $j = p-1$, $b^{i-}_m[x_{m+1}]$ is a Gaussian distribution.

To shorten notations in the following sections, we denote

$$f(x,z) \triangleq \frac{1}{S} \sum_{g=1}^{|L|} \frac{1}{|\Omega_{l_g}|} \mathbb{P}(z|l_g, x) \mathbb{1}_{\Omega_{l_g}}(x). \quad (15)$$

Intuitively, $f(x,z)$ represents the probability of obtaining a given measurement z from robot pose x considering all possible data associations to the $|L|$ landmarks.

Finally, substituting (15) into (14) yields

$$\mathbb{P}(z_{m+1} \mid \gamma_m = i, H^-_{m+1}) \approx \sum_{n=1}^{S} f(x^{n,i}_{m+1}, z_{m+1}). \quad (16)$$

2) $j = p-2$: We now consider calculation of the term $\mathbb{P}(z_{m+2} \mid \gamma_m = i, H^-_{m+2})$ from (6).

Similarly to (11), performing marginalization and chain rule, yields

$$\sum_{g=1}^{|L|} \frac{1}{|\Omega_{l_g}|} \int_{x_{m+2} \in \Omega_{l_g}} \mathbb{P}(z_{m+2} \mid l_g, x_{m+2}) b^{i-}_m[x_{m+2}] dx_{m+2}, \quad (17)$$

where, as in (13), $b^{i-}_m[x_{m+2}] \triangleq \mathbb{P}(x_{m+2} \mid H^-_{m+2}, \gamma_m = i)$.

Performing chain and Bayes rules, marginalizing over x_{m+1} and data association hypotheses for z_{m+1} yields

$$b^{i-}_m[x_{m+2}] = \frac{1}{\eta^i_{m+1}} \sum_{g=1}^{|L|} \frac{1}{|\Omega_{l_g}|} \int_{x_{m+1} \in \Omega_{l_g}} \mathbb{P}(x_{m+2} \mid x_{m+1}, u_{m+1}) \cdot \mathbb{P}(z_{m+1} \mid l_g, x_{m+1}) b^{i-}_m[x_{m+1}] dx_{m+1}, \quad (18)$$

where $\eta^i_{m+1} \triangleq \mathbb{P}(z_{m+1} \mid H^-_{m+2}, \gamma_m = i)$.

In practice, we approximate $b^{i-}_m[x_{m+2}]$ and η^i_{m+1} via sampling, considering S samples from $b^{i-}_m[x_{m+1}]$:

$$b^{i-}_m[x_{m+2}] \approx \sum_{n=1}^{S} \zeta^{n,i}_{m+1} \mathbb{P}(x_{m+2} \mid x^{n,i}_{m+1}, u_{m+1}), \quad (19)$$

and

$$\zeta^{n,i}_{m+1} \triangleq f(x^{n,i}_{m+1}, z_{m+1}) / \hat{\eta}^i_{m+1}, \quad (20)$$

with $\hat{\eta}^i_{m+1} \triangleq \sum_{n=1}^{S} f(x^{n,i}_{m+1}, z_{m+1})$, such that $\sum_{n=1}^{S} \zeta^{n,i}_{m+1} = 1$. Note, the set of samples $\{x^{n,i}_{m+1}\}$ from $b^{i-}_m[x_{m+1}]$ was already obtained from section III-C1.

Further, observe that $b^{i-}_m[x_{m+2}]$ from (19) corresponds to a mixture belief over x_{m+2}, where each of the samples $x^{n,i}_{m+1}$ from the previous step is propagated via the transition model (1). Thus, in context of Sequential Monte Carlo (SMC), this corresponds to the bootstrap particle filter [18], i.e. where the proposal distribution is chosen to be transition model. Yet, here we also account for ambiguous data association aspects. Thus, as we consider in this work Gaussian models (1), $b^{i-}_m[x_{m+2}]$ is a GMM belief with S components, where $\zeta^{n,i}_{m+1}$ is the weight of the nth component.

Note that in (19) and (20) we have $f(x^{n,i}_{m+1}, z_{m+1})$. Instead of calculating it from scratch considering samples $x^{n,i}_{m+1}$ from the propagated belief $b^{i-}_m[x_{m+1}]$, our *key observation* is that it is already available from the calculations for $j = p-1$, see (16).

As before, we approximate the integral in (17) by generating a set $\{x^{n,i}_{m+2}\}^S_{n=1}$ of S samples from the GMM $b^{i-}_m[x_{m+2}]$ from (19), thus approximating $\mathbb{P}(z_{m+2} \mid \gamma_m = i, H^-_{m+2})$ as

$$\frac{1}{S} \sum_{g=1}^{|L|} \frac{1}{|\Omega_{l_g}|} \sum_{n=1}^{S} \mathbb{P}(z_{m+2} \mid l_g, x^{n,i}_{m+2}) \mathbb{1}_{\Omega_{l_g}}(x^{n,,i}_{m+2}), \quad (21)$$

which, recalling the definition (15), can be finally written as

$$\mathbb{P}(z_{m+2} \mid \gamma_m = i, H^-_{m+2}) \approx \sum_{n=1}^{S} f(x^{n,i}_{m+2}, z_{k+2}) \triangleq \hat{\eta}^i_{m+2}. \tag{22}$$

3) General Case: We are now in a position to calculate $\mathbb{P}(z_{m+j} \mid \gamma_m = i, H^-_{m+j})$ for the general case. This is stated in the following Theorem.

Theorem 1: The expression for $\mathbb{P}(z_{m+j} \mid \gamma_m = i, H^-_{m+j})$ from (6) for any $j \in [2, p-1]$ is given by,

$$\mathbb{P}(z_{m+j} \mid \gamma_m = i, H^-_{m+j}) \approx \sum_{n=1}^{S} f(x^{n,i}_{m+j}, z_{m+j}) \triangleq \hat{\eta}^i_{m+j}. \tag{23}$$

A detailed proof via induction is given in the Supplementary [17]. Informally, as in (17), $\mathbb{P}(z_{m+j} \mid \gamma_m = i, H^-_{m+j})$ can be written as

$$\sum_{g=1}^{|L|} \frac{1}{|\Omega_{l_g}|} \int_{x_{m+j} \in \Omega_{l_g}} \mathbb{P}(z_{m+j} \mid l_g, x_{m+j}) b^{i-}_m[x_{m+j}] dx_{m+j}, \tag{24}$$

where $b^{i-}_m[x_{m+j}]$ is a GMM of S components,

$$b^{i-}_m[x_{m+j}] \doteq \sum_{n=1}^{S} \zeta^{n,i}_{m+j-1} \mathbb{P}(x_{m+j} \mid x^{n,i}_{m+j-1}, u_{m+j-1}), \tag{25}$$

where, similar to (20),

$$\zeta^{n,i}_{m+j-1} \triangleq f(x^{n,i}_{m+j-1}, z_{m+j-1}) / \hat{\eta}^i_{m+j-1}, \tag{26}$$

with $\hat{\eta}^i_{m+j-1} \triangleq \sum_{n=1}^{S} f(x^{n,i}_{m+j-1}, z_{m+j-1})$.

The next step is to approximate the integral in (24) via sampling from the GMM $b^{i-}[x_{m+j}]$, which yields (23).

Similarly to Section III-C2, calculation re-use can be performed also for the general case considered here. To see that, note the set of samples $\{x^{n,i}_{m+j-1}\}^S_{n=1}$, as well as the corresponding values $f(x^{n,i}_{m+j-1}, z_{m+j-1})$ and η^i_{m+j-1} are already available to us from calculations performed for the previous step, i.e. for $j-1$, and thus can be conveniently re-used. In contrast, in the naive approach, one would have to re-sample the entire chain from scratch, i.e. starting with $m+1$ and until $m+j-1$.

4) Final Calculation of $\Psi_{m|k}$: Based on (7), the the unnormalized weight $\tilde{w}^i_{m|k}$ is

$$\tilde{w}^i_{m|k} \approx \prod_{j=0}^{p-1} \hat{\eta}^i_{k-j} \cdot w^i_{m|m}. \tag{27}$$

As mentioned in Section III-A, calculating the normalized weight $w^i_{m|k}$ can be done via (8), which requires to first calculate $\tilde{w}^q_{m|k}$ for all hypotheses from time instant m, i.e. $\forall q \in [1, M_m]$.

D. Re-Evaluation in Retrospective of a Specific Data Association Hypothesis

In a similar manner, we can also consider a specific data association hypothesis $\beta_{k-p} = c$ from some past time $k-p$, with $c \in [1, |L|]$, rather than a sequence of data association hypotheses $\gamma_{k-p} \equiv \beta_{1:k-p} = i$ as done above. Note $i \in [1, M_{k-p}]$; without pruning $M_{k-p} = |L|^{k-p}$.

Indeed, by marginalizing over $\gamma_{k-p-1} \equiv \beta_{1:k-p-1}$ we get

$$\mathbb{P}(\beta_{k-p} = c \,|\, H_k) = \sum_{l=0}^{M_{k-p-1}} \mathbb{P}(\beta_{k-p} = c, \beta_{1:k-p-1} = l \,|\, H_k). \tag{28}$$

However, recall $\gamma_{k-p} \equiv \beta_{1:k-p}$, which can assume $M_{k-p} = M_{k-p-1} \cdot |L|$ values. Further, the probability for the ith realization of γ_{k-p} conditioned on H_k, i.e. $w^i_{k-p|k}$ with $i \in [1, M_{k-p}]$, is given by (6).

We now observe the index i designates a combination of some specific realization l of $\gamma_{k-p-1} \equiv \beta_{1:k-p-1}$ and some specific realization r of β_{k-p}. We shall denote these specific realizations for a given i as $i.l$ and $i.r$, respectively (standing for left and right).

Recall that w^i_{k-p} is available for any $i \in [1, M_{k-p}]$ from (6), see Section III-C4. Based on (28), we get

$$\mathbb{P}(\beta_{k-p} = c \,|\, H_k) = \sum_{i=1}^{M_{k-p}} w^i_{k-p|k} \mathbb{1}_{\{c\}}(i.r), \tag{29}$$

i.e. we sum only those realizations of γ_{k-p} that consider the cth data association from time $k-p$.

E. Enhanced Pruning of Hypotheses At Current Time

Another immediate implication of our framework is the ability to enhance hypotheses pruning at current time. In detail, consider we re-evaluated in retrospective hypotheses $i \in [1, M_m]$ from past time $m = k-p$, i.e. $w^i_{k-p|k}$, as discussed in Section III-C4. Then, considering some user-specified pruning threshold th, for any $w^i_{k-p|k} < th$ we can prune also all its descendant hypotheses at time k.

More formally, considering the jth hypothesis at time k with $j \in [1, M_k]$, i.e. $\gamma_k \equiv \beta_{1:k} = j$, we again observe that index j designates a combination of some specific realization of $\beta_{1:k-p}$ and some specific realization of $\beta_{k-p+1:k}$. In a similar fashion to Section III-D, we shall denote these realizations for a given j as $j.l$ and $j.r$, respectively (standing for left and right). Then, for any $w^i_{k-p|k} < th$ we can prune all hypotheses $j \in [1, M_k]$ from current time k that satisfy $j.l = i$. In other words, each hypothesis $j \in [1, M_k]$ is pruned if its ancestor hypothesis from time $k-p$, with index $j.l$ is below the pruning threshold th. The updated hypotheses' weights are therefore

$$\tilde{w}^j_k \doteq w^j_k \cdot \mathbb{1}_{\{j.l\}}(w^{j.l}_{k-p|k} > th). \tag{30}$$

In order to retrieve a valid GMM the weights should be renormalized to sum to one, i.e. $w^j_k \doteq \tilde{w}^j_k / \sum_{j=1}^{M_k} \tilde{w}^j_k$, and the zero-weight (pruned) hypotheses discarded.

We note that, in general, a hypothesis j pruned this way may be above a pruning threshold, i.e. $w^j_k > th$, and thus would not be pruned without re-evaluating its ancestor hypothesis in retrospective, as suggested herein. We demonstrate this aspect in the results section.

F. Computational Complexity Aspects and Algorithm

Calculating $w^i_{m|k}$ involves computation of the unweighted weights $\tilde{w}^q_{m|k}$, using (27), for *each* of the M_m hypotheses from time instant m, and then normalization via (8). For each given

Algorithm 1: $w^i_{m|k}$ Calculation.

Inputs:
H_k: History at time k
$b[X_m]$: GMM belief at time $m = k - p$
i: hypothesis index from time $m = k - p$
for q=1:M_m **do**
Calculate $\tilde{w}^q_{m|k}$ using Algorithm 2.
  end for
$\triangleright$ Normalization via (8)
$w^i_{m|k} = \tilde{w}^i_{m|k} / \sum_{q=1}^{M_m} \tilde{w}^q_{m|k}$
return $w^i_{m|k}$

Algorithm 2: $\tilde{w}^i_{m|k}$ Calculation With Computation Re-use.

Inputs:
H_k: History at time k
$b[X_m]$: GMM belief at time $m = k - p$
q: hypothesis index from time $m = k - p$
for $j = 1 : p$ **do**
  if $j = 1$ **then**
    Sample set $\{x^{n,q}_{m+1}\}^S_{n=1}$ from $b^{q-}_m[x_{m+1}]$, see Section III-C1
  else
    Reuse $\{\zeta^{n,q}_{m+j-1}\}^S_{n=1}$ to form a GMM belief $b^{q-}_m[x_{m+j}]$ from (25)
    Sample set $\{x^{n,q}_{m+j}\}^S_{n=1}$ from $b^{q-}_m[x_{m+j}]$
  end if
  Calculate $\hat{\eta}^q_{m+j}$ using samples $\{x^{n,q}_{m+j}\}^S_{n=1}$ as shown in (23).
  Calculate weights $\{\zeta^{n,q}_{m+j}\}^S_{n=1}$ via (26) for sample set $\{x^{n,q}_{m+j}\}^S_{n=1}$
end for
Calculate and return $\tilde{w}^q_{m|k}$ via (27):

$q \in [1, M_m]$, following (27), this involves calculating $\hat{\eta}^q_{k-j}$ for all $j \in [0, p-1]$, i.e. from $\hat{\eta}^q_{k-p+1} \equiv \hat{\eta}^q_{m+1}$ until $\hat{\eta}^q_k \equiv \hat{\eta}^q_{m+p}$.

In the naïve approach, such calculations are performed from scratch for each $\hat{\eta}^q_{k-j}$. In other words, as described in previous sections, this involves sequentially sampling the beliefs from time instances k until $k - j$, for *each* $j \in [0, p-1]$. Assuming the same number of samples is taken at each time, for a given j, with $j \in [0, p-1]$, this operation involves generating S samples j times. Hence, evaluating (27) is

$$\underbrace{S}_{j=p-1} + \underbrace{2 \cdot S}_{j=p-2} + \ldots + \underbrace{p \cdot S}_{j=0}$$
$$= \frac{(1+p) \cdot p}{2} \cdot S \in \mathcal{O}(p^2 S).$$

As this evaluation has to be performed M_m times, the overall complexity of the naïve approach is $\mathcal{O}(p^2 S M_m)$.

In contrast, our approach uses calculations in a recursive form, therefore we only need N samples for every calculation of η^q_{k-j}. Thus, evaluating (27) is

$$\underbrace{S}_{j=p-1} + \underbrace{S}_{j=p-2} + \ldots + \underbrace{S}_{j=0}$$
$$= p \cdot S \in \mathcal{O}(p \cdot S),$$

and the corresponding overall complexity of our approach is $\mathcal{O}(pSM_m)$, i.e. one order of magnitude smaller in p than the naïve approach.

Algs. 1 and 2 summarize our approach from Section III-C.

IV. Experiments

In this section we examine our proposed algorithm in simulation considering an extremely perceptually aliased environment comprising $|L| = 8$ *identical* spatially scattered static known landmarks. Our simulations are based on the GTSAM library [3] with a Matlab wrapper.

The robot acquires relative pose observations to landmarks during its motion, yet the data association is not assumed to be externally provided, i.e. it is *unknown* which landmark generated each observation. Two scenarios of 5 and 10 time steps and differently scattered landmarks are considered, as shown in Figs. 2(a) and 3(a). In both cases, the robot starts with a uni-modal Gaussian prior belief on its initial location, $b[X_0]$ and performs a pre-defined trajectory.

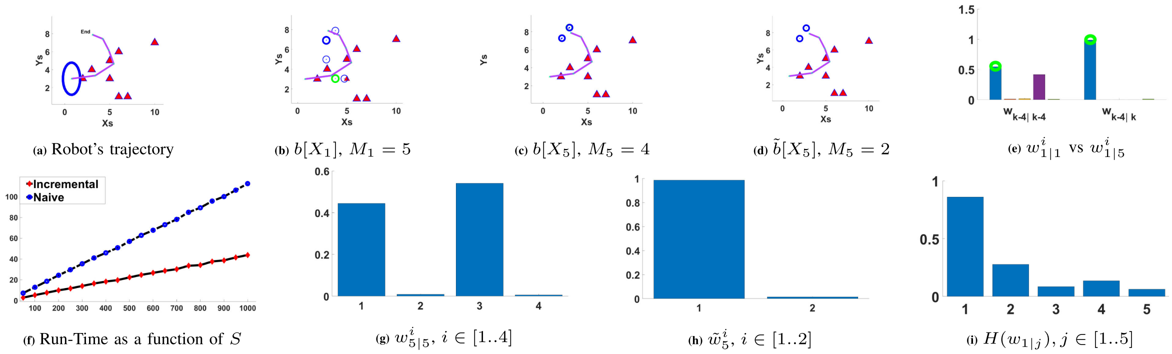
We use a diagonal process covariance matrix Σ_w with standard deviation (std) of position of 0.5 meters and std of orientation of 10^{-3} radians. The measurement covariance matrix Σ_v is also diagonal with position std of 0.48 meters and orientation std of $0.87 \cdot 10^{-2}$ radians.

Figs. 2 and 3 show the results for both scenarios. At each time step k, following [5], [6], we update the belief from the previous time with the performed action and acquired observation with unknown data association, producing $b[X_k]$ from (3). For each case, we show in Figs. 2(b)–(d) and 3(b)–(d) the corresponding posterior GMM belief, shown in blue at specific time instances of interest.

Theoretically, the number of GMM components, grows exponentially according to the recursion $M_k = M_{k-1} \cdot |L|$. However, we prune components with negligible weights and show only the remaining components. Further, similarly to [5], [6], we occasionally merge sufficiently similar components, as shown in Figs. 3(d) and 2(d). We use the notation $\tilde{\Box}$ to denote the belief and components' weights after merging.

Given the above, in both scenarios we consider the ith data association hypothesis from time instant 1, and execute Algorithm 1 to re-evaluate its probability in retrospective, i.e. given current time is k we calculate $w^i_{1|k}$. For each step in our weight update we take a set of $S = 1000$ samples from the GMM created via (25). In our current implementation, samples are taken globally from the entire GMM, where the number of samples per each component is determined according to the GMM weights distribution.

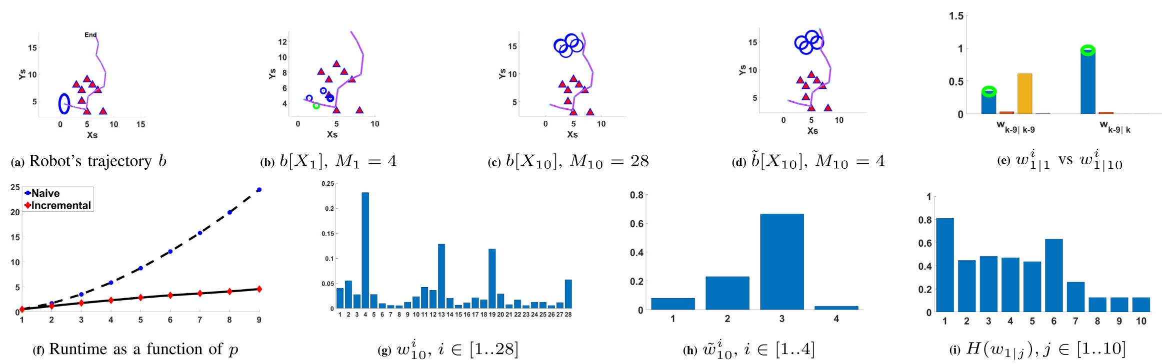
The results for the first and second scenarios are shown, respectively, in Figs. 2(e) and 3(e). In both cases we can see that at $k = 1$ we lack the ability to disambiguate between the hypotheses, however when using the information up to time $k = 5$ and $k = 10$ respectively, we can perform full disambiguation according to the updated weight distribution, via $w_{1|k}$ where $k \in [5, 10]$.



(a) Robot's trajectory (b) $b[X_1]$, $M_1 = 5$ (c) $b[X_5]$, $M_5 = 4$ (d) $\tilde{b}[X_5]$, $M_5 = 2$ (e) $w^i_{1|1}$ vs $w^i_{1|5}$

(f) Run-Time as a function of S (g) $w^i_{5|5}$, $i \in [1..4]$ (h) $\tilde{w}^i_5$, $i \in [1..2]$ (i) $H(w_{1|j})$, $j \in [1..5]$

Fig. 2. First scenario: (a) Five-step ground truth trajectory, an initial prior belief $\mathbb{P}(X_0)$, and known identical landmarks. (b) $b[X_{k=1}]$ is a GMM with five components. We choose this point as our **"re-evaluation"** point, where we test our algorithm. The hypothesis marked in green, states the correct hypothesis generated from the landmark associated to the given measurement. Notice that the **bolder** lines correspond to a higher weight of the hypothesis. (c) GMM belief at $k = 5$; we should note that the number of components doesn't increase from the "re-evaluation" point as a result of pruning. (d) The belief at $k = 5$ after merging, which reduces the number of hypotheses to $M_5 = 2$. (e) The calculated weight distribution at time $k = 1$ given the gathered information up to time $k = 1$ (see (b)), and given information up to time $k = 5$. The green circle marks the weight of the correct hypothesis as shown in (b). (f) Run-Time in seconds of $w^i_{k-4|k}$ calculation vs. number of samples S. The calculation was performed for both the naïve and the proposed incremental approach. (g) The current weight distribution of $b[X_5]$ as shown in (c), via $w^i_{k=5|k=5}$, where $i \in [1..M_5]$ and $M_5 = 4$. (h) Current weight distribution after merging, i.e. $\tilde{b}[X_5]$, where $M_5 = 2$. (i) Entropy distribution of the weights at time $k = 1$ given information at different time points up to time $k = 5$.



(a) Robot's trajectory b (b) $b[X_1]$, $M_1 = 4$ (c) $b[X_{10}]$, $M_{10} = 28$ (d) $\tilde{b}[X_{10}]$, $M_{10} = 4$ (e) $w^i_{1|1}$ vs $w^i_{1|10}$

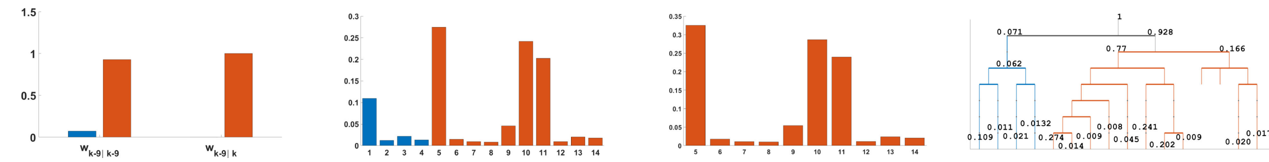
(f) Runtime as a function of p (g) w^i_{10}, $i \in [1..28]$ (h) $\tilde{w}^i_{10}$, $i \in [1..4]$ (i) $H(w_{1|j})$, $j \in [1..10]$

Fig. 3. Second scenario: (a) Ten-step ground truth trajectory, an initial prior belief $\mathbb{P}(X_0)$, and known identical landmarks. (b) The belief at $k = 1$, our **re-eveluation** point, is a GMM with four components. The hypothesis marked in green, states the correct hypothesis generated from the landmark associated to the given measurement. (c) The updated belief at current time, $k = 10$. Due to the ambiguity of the environment, the number of hypotheses increased from $M_1 = 4$ to $M_{10} = 28$ given pruning. (d) The belief at $k = 10$ with the merging effect, reduces the number of hypotheses to $M_{10} = 4$. (e) The calculated weight distribution at time $k = 1$ given the gathered information up to time $k = 1$, and given information up to time $k = 10$. The green circle marks the weight of the correct hypothesis as shown in 3(b). (f) Run-Time in seconds of $w^i_{k-p|k}$ calculation vs. p for the naïve and the proposed incremental approach. (g) The current weight distribution of $b[X_{10}]$, i.e. $w^i_{k=10|k=10}$, where $M_{10} = 28$. (h) Current weight distribution after merging, $\tilde{b}[X_{10}]$, where $M_{10} = 4$. (i) Entropy of the weights at time $k = 1$ given information at different time points up to time $k = 10$.

In addition in the second test case shown in Fig. 3 we see that even after new information has been acquired along the trajectory, we still have an ambiguous setting with or without the merging effect: the corresponding weight distribution of the current belief includes several non-negligible hypotheses at $k = 10$, as shown in Fig. 3(g) and (h). This is in contrast to the first scenario that after merging shows a single none negligible weight as shown in Fig. 2(h). Nevertheless, as shown in Figs. 2(e) and 3(e), our approach to update the weights in retrospective utilizing the information acquired until time k leads to more informative weight distributions. In particular, in both considered scenarios one of the hypotheses' value becomes substantially higher than the rest.

We also report runtime for both the naïve and our incremental approach. For the first scenario we calculated runtime as a function of number of samples $S \in [50, 1000]$, where the result per a number of samples is the statistical average of 20 runs. As shown in Fig. 2(f) the naïve approach calculation grows in a higher rate than the incremental. For the second scenario, as shown in Fig. 3(f), we calculated the runtime as a function of the parameter p, see e.g. (4), where for this analysis the number of samples was fixed at 100. The runtime of the naïve approach shows a growth at a rate of p^2, while the incremental one increases in a linear rate of p, which is in correlation to our calculation in Section III-F.

In addition Figs. 2(i) and 3(i) show the entropy of the weight distribution at the "re-evaluation" point given information at different time instances, i.e. $H(W_{1|j}) = -\sum_{i=1}^{M_1} w^i_{1|j} \cdot \log(w^i_{1|j})$ where $j \in [1, k]$. Both test cases show that as we use more information at the "re-evaluation" point, the level of uncertainty



(a) $w^i_{1|1}$ vs $w^i_{1|10}$ (b) w^i_{10}, $i \in [1..14]$ (c) $\tilde{w}^i_{10}$, $i \in [5..14]$ (d) GMM weights spanning tree.

Fig. 4. (a) The calculated weight distribution at time $k = 1$ given the gathered information up to time $k = 1$, and given information up to time $k = 10$. (b) The weight distribution at $k = 10$. The descendants of w^1_1 and w^2_1 are marked in blue and orange, respectively. In total we have $M_{10} = 14$ hypotheses (GMM components). (c) The updated weight distribution at $k = 10$, where all descendants of w^1_1 have been discarded after re-evaluation of the weights at $k = 1$. The number of hypotheses reduces to 10. (d) The entire hypothesis tree describing the evolution of hypotheses (GMM belief components) from the initial time and until the current time $k = 10$. Blue and orange colors represent descendant hypotheses of w^1_1 and w^2_1, respectively.

can reduce, although we should mention that this is not guaranteed, since it is conditioned on the landmarks setting, and the randomized samples values.

Enhanced Pruning of Hypotheses at Current Time: In a similar simulation setting comprising eight identical landmarks and a trajectory of 10 time steps we examined the enhanced pruning of hypotheses at *current* time after re-evaluation in retrospective of past hypotheses (at $k = 1$), as discussed in Section III-E. Results are shown in Fig. 4.

Fig. 4(a) shows that at time $k = 1$ there are two un-pruned hypotheses, i.e. whose weights are higher value than the pruning threshold $th = 0.005$. After we perform re-evaluation, hypotheses $w^1_{1|10}$ was pruned, and we remain with a single hypotheses $w^1_{2|10}$ marked in orange. As discussed in Section III-E, all the descendant hypotheses of $w^1_{1|10}$ can now be discarded. These hypotheses are shown in blue color in the GMM "hypotheses tree" in Fig. 4(d). More in detail, Fig. 4(b) shows the weights of the original GMM belief at $k = 10$, with $M_{10} = 14$ components: 4 descendant hypotheses of w^1_1 and 10 of w^2_1, marked in blue and orange, respectively. Note all components are *above* the pruning threshold th. Fig. 4(c) shows the reduced GMM after pruning all the blue hypotheses (descendants of $w^1_{1|10}$) and re-normalizing the weights. Thus, leveraging the proposed concept the number of components reduces to 10.

V. Conclusion

We presented an approach to update probabilities of externally-specified hypotheses from some past time with information obtained since then and until the current time. Our approach is particularly of interest in the context of robust perception and autonomous navigation in ambiguous and perceptually aliased scenarios, which necessitate reasoning about data association hypotheses and thus to maintain mixture distributions such as GMM. In addition we developed an incremental form for calculation re-use, as opposed to a naïve approach that performs recalculation in each step. Another direct consequence of our approach is enhanced pruning of hypotheses at current time, leveraging the updated weights of corresponding ancestor hypotheses given information thus far. Our simulation shows that re-evaluation of a past time in a highly aliased setting can assist with hypotheses disambiguation both in past and current time, and with our ability to discard an entire branch(es) from the GMM hypothesis tree. In future research we aim to study the performance of our approach in real world settings, explore how adding landmarks to the state (i.e. a high dimensional setting) effects the computational complexity, and investigate how it can be incorporated within the planning phase.

References

[1] C. Cadena *et al.*, "Simultaneous localization and mapping: Present, future, and the robust-perception age," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, Dec. 2016.

[2] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G2o: A general framework for graph optimization," in *Proc. IEEE Int. Conf. Robot. Automat.*, Shanghai, China, 2011, pp. 3607–3613.

[3] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," Georgia Inst. Technol., Tech. Rep. GT-RIM-CP&R-2012-002, Sep. 2012.

[4] S. Agarwal, K. Mierle, and Others, "Ceres solver," Google, Tech. Rep., 2016. [Online]. Available: http://ceres-solver.org

[5] S. Pathak, A. Thomas, and V. Indelman, "A unified framework for data association aware robust belief space planning and perception," *Int. J. Robot. Res.*, vol. 32, no. 2-3, pp. 287–315, 2018.

[6] V. Tchuiev, Y. Feldman, and V. Indelman, "Data association aware semantic mapping and localization via a viewpoint-dependent classifier model," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2019, pp. 7742–7749.

[7] M. Hsiao and M. Kaess, "MH-iSAM2: Multi-hypothesis isam using bayes tree and hypo-tree," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2019, pp. 1274–1280.

[8] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Multi-target tracking using joint probabilistic data association," in *Proc. 19th IEEE Conf. Decis. Control*, 1980, pp. 807–812.

[9] N. Sünderhauf and P. Protzel, "Switchable constraints for robust pose graph SLAM," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2012, pp. 1879–1884.

[10] E. Olson and P. Agarwal, "Inference on networks of mixtures for robust robot mapping," *Int. J. Robot. Res.*, vol. 32, no. 7, pp. 826–840, 2013.

[11] L. Carlone, A. Censi, and F. Dellaert, "Selecting good measurements via l1 relaxation: A convex approach for robust estimation over graphs," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2014, pp. 2667–2674.

[12] V. Indelman, E. Nelson, J. Dong, N. Michael, and F. Dellaert, "Incremental distributed inference from arbitrary poses and unknown data association: Using collaborating robots to establish a common reference," *IEEE Control Syst. Mag.*, vol. 36, no. 2, pp. 41–74, Apr. 2016.

[13] L. Wong, L. P. Kaelbling, and T. Lozano-Pérez, "Data association for semantic world modeling from partial views," *Int. Symp. Robot. Res.*, vol. 34, no. 7, pp. 1064–1082, 2015.

[14] H. Yang, P. Antonante, V. Tzoumas, and L. Carlone, "Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 1127–1134, Apr. 2020.

[15] P. Antonante, V. Tzoumas, H. Yang, and L. Carlone, "Outlier-robust estimation: Hardness, minimally-tuned algorithms, and applications," 2020, *arXiv:2007.15109*.

[16] D. Fourie, J. Leonard, and M. Kaess, "A nonparametric belief solution to the Bayes tree," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2016, pp. 2189–2196.

[17] O. Shelly and V. Indelman, "Hypotheses disambiguation in retrospective," Technion - Israel Inst. Technol., Tech. Rep., 2022. [Online]. Available: https://tinyurl.com/j36p864h

[18] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *Proc. IEE F (Radar Signal Process.)*, 1993, vol. 140, no. 2, pp. 107–113.