

Bayesian Multi-Hypothesis Scan Matching

Edmund Brekke and Mandar Chitre
ARL, Tropical Marine Science Institute
National University of Singapore
14 Kent Ridge Road
Singapore 119222
Email: {edmund, mandar}@arl.nus.edu.sg

Abstract—This paper proposes a multi-hypothesis solution to the simplified problem of simultaneous localization and mapping (SLAM) that arises when only two measurement frames are available. The proposed solution calculates hypothesis probabilities according to modeling based on standard multitarget tracking (MTT). State estimation is carried out by a hybrid technique consisting of extended Kalman filtering (EKF) and natural gradient (NG) optimization. The search for promising candidate hypotheses is carried out by Bron & Kerbosh’ clique detection algorithm. Both Monte-Carlo simulations and implementation on real-world sonar data show that the proposed approach has desirable robustness properties.

I. INTRODUCTION

A key task in the navigation of autonomous vehicles is to estimate the motion of the vehicle with respect to its surrounding environment. This naturally leads to the concept of simultaneous localization and mapping (SLAM), in which the vehicle builds a map of the environment while simultaneously estimating its own position in this map. SLAM is typically formulated according to a feature-based parametrization (FB-SLAM), where estimation is done by processing sets of point measurements extracted from images obtained by a radar, sonar or laser scanner.

In this paper we address the special case of FB-SLAM which arises when only two consecutive data frames, and corresponding measurement sets, are available. We refer to this problem as feature-based scan-matching (FBSM). An example application of FBSM is initialization of recursive SLAM methods. Furthermore, we believe that it is instructive to study this simplified problem in close detail, before attempting to pursue Bayes-optimal solutions to the full SLAM problem which comprises several data frames.

FB-SLAM is in some ways related to the problem of multi-target tracking (MTT). The difference is, crudely speaking, that in FB-SLAM the sensor moves and the targets are stationary, while in MTT the target moves while the sensor is stationary. Both problems are naturally decomposed into two subproblems: *data association* and *state estimation*. The former problem concerns establishing which features in consecutive scans originate from the same physical objects. The latter problem, when viewed from the perspective of SLAM, concerns estimating the motion of the vehicle and the map of landmarks, conditional on such data association hypotheses.

Several solutions to data association in SLAM have been inspired by corresponding developments in MTT. This can be said about the multi-dimensional assignment approach of [1], the probabilistic multi-hypothesis tracking (PMHT) based

SLAM of [2], FastSLAM [3] and various methods based on finite set statistics (FISST) [4].

Both in MTT and in SLAM, the most reasonable benchmark methods are arguably multi-hypothesis approaches. The multiple hypothesis tracker (MHT) [5] is widely regarded as optimal in MTT [6], while joint branch and bound (JCBB) [7] is considered a gold standard in SLAM [8]. Both these methods have in common that they consider several association hypotheses with the aim of choosing the best one, but they differ in some important aspects. The MHT calculates posterior probabilities of all tentative association hypotheses in a Bayesian manner, while JCBB attempts to find the best hypothesis by using the number of feasible correspondences as a score function.

The purpose of this paper is to present a solution similar to the MHT for the FBSM problem, and thus to make a first step towards extending the presumably optimal formalism of the MHT to SLAM in general. This entails four contributions. First, we derive a formal solution to the FBSM problem based on standard MTT modeling. Second, we present a closed form approximation for the probabilities of association hypotheses. Third, we suggest that the extended Kalman filter (EKF) should be supplemented with a non-linear optimization technique such as Amari’s natural gradient (NG) [9] in order to achieve sufficient accuracy for adequate evaluation of hypothesis probabilities. Fourth, we devise an efficient strategy for exploring the hypothesis space. This strategy is based on the Bron-Kerbosch method for clique-detection in graph theory [10]. We therefore refer to the overall FBSM method as clique-detection scan-matching (CDSM).

This paper is organized as follows. After some brief notational comments in Section II, the conceptual framework of the FBSM problem and its multi-hypothesis solution is outlined in Section III. The methodology of CDSM is explained in Sections IV, V and VI, which concern hypothesis probability evaluation, pose estimation and hypothesis search techniques, respectively. Simulation results are presented in Section VII, while results on real sonar data are presented in Section VIII. Finally, a brief conclusion is provided in Section IX.

II. NOTATION AND TERMINOLOGY

Let the 2-dimensional rotation matrix be given by

$$R(\psi) = \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix}. \quad (1)$$

Landmark positions (a.k.a. targets) are generally denoted by x . Vehicle *position* is denoted by $\rho_k = [x_k, y_k]^T$, while

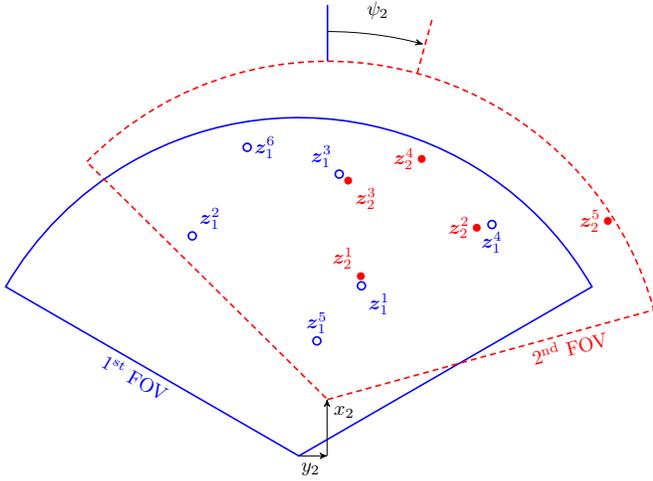


Fig. 1: Illustration of the scan-matching problem. Our aim is to estimate the vehicle pose $\mathbf{p}_2 = [x_2, y_2, \psi_2]^T$ from the measurement sets \mathbf{Z}_1 and \mathbf{Z}_2 . In this illustrative example a plausible association hypothesis would consist of the correspondences (z_1^1, z_2^1) , (z_1^4, z_2^2) and (z_1^3, z_2^3) .

vehicle pose is denoted by $\mathbf{p}_k = [x_k, y_k, \psi_k]^T$ (cf. Fig. 1). By $\mathcal{N}(\mathbf{a}; \mathbf{b}, \Sigma)$ we mean the Gaussian probability density function (pdf) $\mathcal{N}(\mathbf{b}, \Sigma)$ evaluated at \mathbf{a} .

A *correspondence* consists of two measurements from different time steps. An *association hypothesis* is a set of correspondences (see Section III-D). A *minimal sample set* (MSS) is an association hypothesis of cardinality two, i.e., with only two correspondences.

III. CONCEPTUAL FRAMEWORK

In this section we phrase the FBSM problem, i.e., the 2-frame SLAM problem, according to a classical Bayesian framework along the lines of [1] and [5].

A. Kinematic prior

We assume that there exist a set $\mathbf{X} = \{\mathbf{x}^1, \dots, \mathbf{x}^n\}$ of n stationary landmarks, where $\mathbf{x}^i = [\rho_x^i, \rho_y^i]^T$, and both \mathbf{x}^i and n are unknown. The landmarks are *a priori* independently and identically distributed (i.i.d.) according to uniform distribution over a region \mathcal{S} :

$$f(\mathbf{x}^i) = \frac{1}{V(\mathcal{S})} \chi_{\mathcal{S}}(\mathbf{x}^i). \quad (2)$$

Furthermore, the landmarks are assumed *a priori* independent of the vehicle pose \mathbf{p}_k . We base our work on the standard motion model

$$\mathbf{p}_2 = \mathbf{p}_1 \oplus \mathbf{q} + \mathbf{w}, \quad \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \quad (3)$$

where the displacement \mathbf{q} depends on the velocity of the vehicle, \mathbf{w} is plant noise, and the compounding operator \oplus is defined as in [11]. We employ the robocentric assumption that \mathbf{p}_1 is perfectly known ($= \mathbf{0}_{3 \times 1}$), and that the prior knowledge about \mathbf{q} can be represented by $\mathcal{N}(\boldsymbol{\mu}_q, \mathbf{P}_q)$. The prior knowledge about \mathbf{p}_2 can then be represented by

$$f(\mathbf{p}_2) = \mathcal{N}(\mathbf{p}_2; \boldsymbol{\mu}_q, \mathbf{P}) \quad \text{where } \mathbf{P} = \mathbf{Q} + \mathbf{P}_q. \quad (4)$$

B. Measurement model

At each time step a measurement set $\mathbf{Z}_k = \{z_k^1, \dots, z_k^{m_k}\}$ is registered. Any measurement z_k^j may originate from clutter or from one of the landmarks in \mathbf{X} . In the former case its spatial pdf is uniform over the sensor's field of view:

$$f(z_k^j | \text{Not from landmark}) = 1/V. \quad (5)$$

In the latter case its spatial pdf is modeled as a Gaussian

$$f(z_k^j | \mathbf{x}^i, \mathbf{p}_k) = \mathcal{N}(z_k^j; \mathbf{h}(\mathbf{x}^i, \mathbf{p}_k), \mathbf{R}) \quad (6)$$

where

$$\mathbf{h}(\mathbf{x}^i, \mathbf{p}_k) = \mathbf{f}(\mathbf{g}(\mathbf{x}^i, \mathbf{p}_k)) \quad (7)$$

$$\mathbf{f} \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} \sqrt{x^2 + y^2} \\ \text{atan2}(y, x) \end{bmatrix} \quad (8)$$

$$\mathbf{g}(\mathbf{x}^i, \mathbf{p}_k) = \mathbf{R}(-\psi_k)(\mathbf{x}^i - \boldsymbol{\rho}_k). \quad (9)$$

The matrix \mathbf{R} is related to the sensor resolution [12]. The measurements in \mathbf{Z}_k are independent when conditioned on \mathbf{X} and \mathbf{p}_k .

C. Cardinality models

In addition to the randomness of the kinematic quantities discussed above, we must also specify the random nature of the *numbers* of landmarks, landmark-originating measurements and clutter measurements. We want to make minimal assumptions regarding these numbers. Therefore, we model both the number of landmarks n and the number of clutter points ϕ_k as uniformly distributed with upper limits N and M , respectively:

$$\Pr(n) = \begin{cases} \frac{1}{N+1} & \text{if } n \in \{0, \dots, N\} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

$$\Pr(\phi_k) = \begin{cases} \frac{1}{M+1} & \text{if } \phi_k \in \{0, \dots, M\} \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

Furthermore, we assume that all landmarks are detected with unity probability, so that n of the m_k measurements in \mathbf{Z}_k are generated by landmarks and $m_k = n + \phi_k$.

Regarding the assumptions (10) and (11) it should be noted that the numbers N and M never enter the Bayesian machinery, since the posterior only is defined up to proportionality. The assumption of unity detection probability means that any landmark that fails to be detected twice will be considered as clutter. This is reasonable, since such a landmark provides no useful information in FBSM.

D. Association hypotheses and the MHT framework

We develop the concept of association hypotheses through 4 successive refinements of the outcome space. Based on the terminologies used in [5] and [13], we call these number event, configuration event, data-to-data hypothesis and landmark-to-data hypothesis. The *number event* γ concerns how many measurements are associated with landmarks and with clutter. The *configuration event* τ_k at time k concerns how the set of measurements in \mathbf{Z}_k is partitioned into subsets associated with each of these two sources. The *data-to-data hypothesis* θ concerns how measurements in \mathbf{Z}_2 are associated to measurements in \mathbf{Z}_1 . The *landmark-to-data hypothesis* ω concerns how measurements are generated by specific landmarks. Although the fourth refinement is required for a rigorous Bayesian analysis, it is the third refinement which matters from a

practical perspective. Given a set of two targets and two successive measurements z_1^1 and z_2^1 , we do not distinguish between the possibility that target 1 generated both z_1^1 and z_2^1 , and the possibility that target 2 generated both z_1^1 and z_2^1 . What matters is whether the two measurements were generated by the same target, or by different targets. Therefore, we will in the sequel simply refer to data-to-data hypotheses as *association hypotheses*.

We define a landmark-to-data hypothesis (under the assumptions of Section III-B) as a mapping $\omega : \{1, \dots, n\} \rightarrow \{0, \dots, m_1\} \times \{0, \dots, m_2\}$ such that $\omega_l(s) = \omega_l(t) \Rightarrow s = t$ for any $l \in \{1, 2\}$. Both number events, configuration events and association hypotheses can be rigorously defined as equivalence classes of such mappings. In particular, an association hypothesis is an equivalence class of permutation-equivalent landmark-to-data hypotheses [6]. From this it follows that an association hypothesis also can be viewed as a set of correspondences. A landmark-to-data hypothesis can also be viewed as an $n \times 2$ -matrix, where each column represents a correspondence.

For a given landmark-to-data hypothesis ω we use the following notations. The measurement associated to landmark i at time step k under ω is denoted $\mathbf{z}_k^{\omega(i)}$. By $\mathbf{z}^{\omega(i)}$ we denote the vector $[(\mathbf{z}_1^{\omega(i)})^\top, (\mathbf{z}_2^{\omega(i)})^\top]^\top \in \mathbb{R}^4$, while \mathbf{z}^ω denotes the vector $[(\mathbf{z}_1^{\omega(1)})^\top, (\mathbf{z}_1^{\omega(2)})^\top, \dots, (\mathbf{z}_2^{\omega(n)})^\top]^\top \in \mathbb{R}^{4n}$, i.e., the stacked vector of all measurements that originate from landmarks according to ω . The notation \mathbf{x}^ω signifies a stacked vector containing all the landmarks $[(\mathbf{x}^1)^\top, \dots, (\mathbf{x}^n)^\top]^\top$.

Since we generally care more about data-to-data hypotheses than about landmark-to-data hypotheses, it is often desirable to extend this notation to association hypotheses as well. Unfortunately, expressions such as $f(\mathbf{z}^{\theta(i)}|\mathbf{x}^i)$ are not well-defined since θ does not specify any order on the landmarks. Expressions such as $\prod_{i=1}^n \int f(\mathbf{z}^{\theta(i)}|\mathbf{x}^i)f(\mathbf{x}^i)d\mathbf{x}^i$ are somewhat safer to use since the value of such an expression will remain the same if θ is substituted with any permutation-equivalent ω . We also use expressions such as $f(\mathbf{p}_2, \mathbf{x}^\theta)$ in this paper, even when no integration is performed. This is to be understood in the following sense: $\mathbf{x}^\theta = [(\mathbf{x}^1)^\top, \dots, (\mathbf{x}^n)^\top]^\top$ is a vector containing n landmarks, where each landmark corresponds to a correspondence in θ . Although this notation technically speaking is ill-defined, a notation such as $\mathbf{x}_{2|2}^\theta$, that is, a state estimate of all landmarks involved in θ , carries a unique meaning, since any permutation-equivalent ω would yield the same landmark estimates for the respective correspondences insofar as the same estimation procedure is used and the prior i.i.d. landmark assumption holds.

In order to illustrate the entities defined in this section, let us refer to the example displayed in Figure 1. Here, the correct number event is $\gamma_{\text{true}} = (n = 3, \phi_1 = 3, \phi_2 = 2)$. The correct association hypothesis θ_{true} corresponds to the following set of correspondences:

$$\left\{ \left[\begin{array}{c} 1 \\ 1 \end{array} \right], \left[\begin{array}{c} 3 \\ 3 \end{array} \right], \left[\begin{array}{c} 4 \\ 2 \end{array} \right] \right\}. \quad (12)$$

There are 6 landmark-to-data hypotheses which agree with θ_{true} . Thus, θ_{true} is, rigorously speaking, the equivalence class

$$\left\{ \left[\begin{array}{ccc} 1 & 3 & 4 \\ 1 & 3 & 2 \end{array} \right], \dots, \left[\begin{array}{ccc} 4 & 3 & 1 \\ 2 & 3 & 1 \end{array} \right] \right\}. \quad (13)$$

Having established an MHT formalism for the FBSM problem, it is relatively straightforward to develop expressions for the hypothesis probabilities $\Pr(\theta | \mathbf{Z}_{1:2})$. The simplest possible approach is to use Bayes' rule to obtain¹

$$\Pr(\theta | \mathbf{Z}_{1:2}) = \frac{1}{c} f(\mathbf{Z}_{1:2} | \theta) \Pr(\theta). \quad (14)$$

In accordance with the above discussion, we evaluate the prior hypothesis probability $\Pr(\theta)$ progressively in terms of the number event γ , the configurations τ_1 and τ_2 , and the association hypothesis θ . Based on (10) and (11) we find the probability of γ as

$$\begin{aligned} \Pr(\gamma) &= \Pr(n) \Pr(\phi_1) \Pr(\phi_2) \\ &= \frac{1}{N+1} \frac{1}{(M+1)^2}. \end{aligned} \quad (15)$$

The probability of τ_1 conditional on γ is

$$\Pr(\tau_1 | \gamma) = \frac{n! \phi_1!}{m_1!} \quad (16)$$

while the probability of τ_2 conditional on τ_1 (and thus also conditional on γ) is

$$\Pr(\tau_2 | \tau_1, \gamma) = \frac{n! \phi_2!}{m_2!}. \quad (17)$$

Given both τ_1 and τ_2 , we have $n!$ possible association hypotheses. Treating all of these as equally *a priori* probable yields

$$\Pr(\theta | \tau_1, \tau_2) = \frac{1}{n!}. \quad (18)$$

Thus, the prior probability of θ becomes

$$\begin{aligned} \Pr(\theta) &= \Pr(\theta | \tau_1, \tau_2, \gamma) \Pr(\tau_2 | \tau_1, \gamma) \Pr(\tau_1 | \gamma) \\ &= \frac{n! \phi_1! \phi_2!}{(N+1)(M+1)^2 m_1! m_2!} \propto n! \phi_1! \phi_2!. \end{aligned} \quad (19)$$

The kinematic term $f(\mathbf{Z}_{1:2} | \theta)$ is found according to the total probability theorem as

$$f(\mathbf{Z}_{1:2} | \theta) = \int f(\mathbf{Z}_{1:2} | \boldsymbol{\xi}^\theta, \theta) f(\boldsymbol{\xi}^\theta | \theta) d\boldsymbol{\xi}^\theta. \quad (20)$$

Here $\boldsymbol{\xi}^\theta$ is a joint state vector which contains both the vehicle displacement \mathbf{p}_2 as well as landmark states corresponding to the correspondences in θ :

$$\boldsymbol{\xi}^\theta = [\mathbf{p}_2^\top, (\mathbf{x}^\theta)^\top]^\top = [\mathbf{p}_2^\top, (\mathbf{x}^1)^\top, \dots, (\mathbf{x}^n)^\top]^\top. \quad (21)$$

The pdf's involved in (20) can be written as

$$f(\mathbf{Z}_{1:2} | \boldsymbol{\xi}^\theta, \theta) = \frac{\prod_{i=1}^n f(\mathbf{z}_1^{\theta(i)} | \mathbf{x}^i, \mathbf{0}_3) f(\mathbf{z}_2^{\theta(i)} | \mathbf{x}^i, \mathbf{p}_2)}{V^{\phi_1 + \phi_2}} \quad (22)$$

$$f(\boldsymbol{\xi}^\theta | \theta) = f(\mathbf{p}_2) \prod_{i=1}^n f(\mathbf{x}^i). \quad (23)$$

The expression for $f(\mathbf{Z}_{1:2} | \theta)$ is well-defined since all landmarks are being integrated out, and since $f(\boldsymbol{\xi}^\theta | \theta)$ is symmetric in the landmarks.

¹A more precise approach that would not depend on abuse of notation would be to find $\Pr(\theta | \mathbf{Z}_{1:2}) = \sum_{\omega \in \theta} \Pr(\omega | \mathbf{Z}_{1:2})$ where each $\Pr(\omega | \mathbf{Z}_{1:2})$ is obtained using Bayes' rule.

IV. EVALUATION OF HYPOTHESIS PROBABILITIES

In this section we present the key contribution of this paper. The posterior probability of the association hypothesis θ is given by

$$\Pr(\theta | \mathbf{Z}_{1:2}) = \frac{1}{c} \frac{n! \phi_1! \phi_2!}{V^{\phi_1 + \phi_2}} a^\theta \quad (24)$$

where

$$a^\theta = \int f(\mathbf{p}_2) \prod_{t=1}^n \left(\int f(\mathbf{z}_1^{\theta_1(t)} | \mathbf{x}^t) f(\mathbf{z}_2^{\theta_2(t)} | \mathbf{x}^t, \mathbf{p}_2) f(\mathbf{x}^t) d\mathbf{x}^t \right) d\mathbf{p}_2. \quad (25)$$

Furthermore, we claim that it is possible to approximate a^θ reasonably well by a closed-form expression. In order to present this approximation we convert the measurements to Cartesian coordinates along the lines of [14]. We define the converted measurements as $\mathbf{y}_k^i = \mathbf{f}^{-1}(\mathbf{z}_k^i)$, that is, according to

$$\mathbf{y}_k^i = \begin{bmatrix} r \cos \vartheta \\ r \sin \vartheta \end{bmatrix} \text{ with } r, \vartheta \text{ given by } \mathbf{z}_k^i = \begin{bmatrix} r \\ \vartheta \end{bmatrix}. \quad (26)$$

The covariances corresponding to \mathbf{y}_k^i and \mathbf{z}_k^i are

$$\mathbf{Y}_k^i = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \text{ and } \mathbf{R} = \begin{bmatrix} \sigma_r^2 & 0 \\ 0 & \sigma_\vartheta^2 \end{bmatrix} \quad (27)$$

where (as a first order approximation)

$$\begin{aligned} Y_{11} &= r^2 \sigma_\vartheta^2 \sin^2(\vartheta) + \sigma_r^2 \cos^2(\vartheta) \\ Y_{22} &= r^2 \sigma_\vartheta^2 \cos^2(\vartheta) + \sigma_r^2 \sin^2(\vartheta) \\ Y_{12} &= Y_{21} = (\sigma_r^2 - r^2 \sigma_\vartheta^2) \sin(\vartheta) \cos(\vartheta). \end{aligned} \quad (28)$$

We can then approximate a^θ as

$$a^\theta \approx \begin{cases} \frac{(2\pi)^{3/2} \mathcal{N}(\hat{\mathbf{p}}_{2|2}^\theta; \boldsymbol{\mu}_q, \mathbf{P})}{(V(S)|\mathbf{R}|)^n \sqrt{|\mathbf{J}^\theta(\hat{\mathbf{p}}_{2|2}^\theta)|}} \times \\ \prod_{i=1}^n \sqrt{|\mathbf{Y}_1^{\theta_1(i)}| |\mathbf{Y}_2^{\theta_2(i)}|} s^{\theta(i)}(\hat{\mathbf{p}}_{2|2}^\theta) & n > 0 \\ 1 & n = 0 \end{cases} \quad (29)$$

where

$$s^{\theta(i)}(\mathbf{p}_2) = \mathcal{N}(\mathbf{y}_2^{\theta_2(i)}; \mathbf{A}\mathbf{y}_1^{\theta_1(i)} + \mathbf{b}, \mathbf{Y}_2^{\theta_2(i)} + \mathbf{A}\mathbf{Y}_1^{\theta_1(i)}\mathbf{A}^\top) \quad (30)$$

and where

$$\mathbf{A} = \mathbf{R}(-\psi_2) \quad \mathbf{b} = -\mathbf{R}(-\psi_2)\boldsymbol{\rho}_2. \quad (31)$$

Notice that both \mathbf{A} and \mathbf{b} depend on \mathbf{p}_2 . The vector $\hat{\mathbf{p}}_{2|2}^\theta$ is an MAP estimate of \mathbf{p}_2 conditional on θ , which is found using EKF- and NG-techniques as explained in Section V. Finally, $\mathbf{J}^\theta(\hat{\mathbf{p}}_{2|2}^\theta)$ is an information matrix which describes the posterior density of \mathbf{p}_2 . It is found as a sum of information matrices for the correspondences involved in the association hypothesis θ :

$$\mathbf{J}^\theta(\mathbf{p}_2) = \mathbf{P}^{-1} + \sum_{i \in \theta} \mathbf{J}^{\theta(i)}(\mathbf{p}_2). \quad (32)$$

By defining

$$\begin{aligned} \boldsymbol{\nu}^{\theta(i)} &= \mathbf{y}_2^{\theta_2(i)} - \mathbf{A}\mathbf{y}_1^{\theta_1(i)} - \mathbf{b} \\ \boldsymbol{\Sigma}^{\theta(i)} &= \mathbf{Y}_2^{\theta_2(i)} + \mathbf{A}\mathbf{Y}_1^{\theta_1(i)}\mathbf{A}^\top \end{aligned} \quad (33)$$

we can, by means of a result in [15], find the correspondence-conditional information matrices $\mathbf{J}^{\theta(i)}$ as

$$\begin{aligned} \mathbf{J}^{\theta(i)}(\mathbf{p}_2) &= \left(\mathcal{D}_{\mathbf{p}_2} \boldsymbol{\nu}^{\theta(i)} \right)^\top (\boldsymbol{\Sigma}^{\theta(i)})^{-1} \mathcal{D}_{\mathbf{p}_2} \boldsymbol{\nu}^{\theta(i)} \\ &+ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} \text{tr} \left((\boldsymbol{\Sigma}^{\theta(i)})^{-1} \mathbf{U}^{\theta(i)} (\boldsymbol{\Sigma}^{\theta(i)})^{-1} \mathbf{U}^{\theta(i)} \right) \end{bmatrix} \end{aligned} \quad (34)$$

where

$$\mathcal{D}_{\mathbf{p}_2} \boldsymbol{\nu}^{\theta(i)} = \left[\mathbf{A}, \begin{bmatrix} -\sin \psi & \cos \psi \\ -\cos \psi & -\sin \psi \end{bmatrix} (\boldsymbol{\rho}_2 - \mathbf{y}_1^{\theta_1(i)}) \right],$$

and the matrix

$$\mathbf{U}^{\theta(i)} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \quad (35)$$

is given according to

$$\begin{aligned} u_{11} &= -2c_{11} \sin \psi + 2c_{12} \cos \psi \\ u_{21} &= -(c_{12} + c_{21}) \sin \psi + (c_{22} - c_{11}) \cos \psi \\ u_{12} &= -(c_{12} + c_{21}) \sin \psi + (c_{22} - c_{11}) \cos \psi \\ u_{22} &= -2c_{22} \sin \psi - 2c_{21} \cos \psi \end{aligned} \quad (36)$$

where c_{11} , c_{12} , c_{21} and c_{22} are given by

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = \mathbf{A}\mathbf{Y}_1^{\theta_1(i)}. \quad (37)$$

A full derivation of these results is beyond the scope of this paper, and will be covered elsewhere. With regard to the general expression (24) it should be obvious how $\Pr(\theta | \mathbf{Z}_{1:2})$ can be split into the factors $\Pr(\mathbf{Z}_{1:2} | \theta)$ and $\Pr(\theta)$ as given in the previous section. With regard to the closed-form approximation of a^θ , our solution is inspired by observing that the hypothesis-conditioned posteriors $f(\boldsymbol{\xi}^\theta | \mathbf{Z}_{1:2}, \theta)$ tend to be close to Gaussian, see Figure 2. For both the inner and the outer integrals in (25) we approximate each integrand as a Gaussian multiplied with the ratio between the peak of the true integrand and the peak of the approximating Gaussian. This makes the integration trivial, and we are left with the peak ratios.

V. STATE ESTIMATION

Under the given assumptions, the pdf $f(\boldsymbol{\xi}^\theta | \theta, \mathbf{Z}_1)$ is approximately equal to $\mathcal{N}(\boldsymbol{\xi}; \boldsymbol{\xi}_{1|1}^\theta, \mathbf{P}_{1|1}^\theta)$ where

$$\boldsymbol{\xi}_{1|1}^\theta = \begin{bmatrix} \boldsymbol{\mu}_q \\ \mathbf{y}_1^{\theta(1)} \\ \vdots \\ \mathbf{y}_1^{\theta(n)} \end{bmatrix}, \quad \mathbf{P}_{1|1}^\theta = \begin{bmatrix} \mathbf{P} & & & \\ & \mathbf{Y}_1^{\theta(1)} & & \\ & & \ddots & \\ & & & \mathbf{Y}_1^{\theta(n)} \end{bmatrix}.$$

The posterior pdf $f(\boldsymbol{\xi}^\theta | \theta, \mathbf{Z}_{1:2})$ can also be approximated by a Gaussian. In order to find the moments of this Gaussian we propose a two-stage estimation strategy. First we obtain an estimate $\boldsymbol{\xi}_{2|2}^\theta$ of the full state vector $\boldsymbol{\xi}^\theta$ by means of an EKF. Then we obtain an improved estimate $\hat{\mathbf{p}}_{2|2}^\theta$ of the pose vector by means of a technique know as the natural gradient [9].

This second step is warranted because the EKF alone is not sufficiently accurate for adequate evaluation of the hypothesis probabilities $\Pr(\theta | \mathbf{Z}_{1:2})$. It is well known that the EKF tends to exhibit inconsistent behavior: The actual estimation error of the EKF may be substantially larger than indicated by

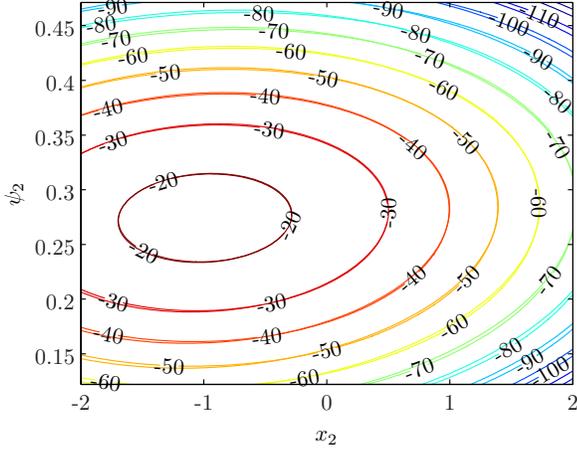


Fig. 2: Contours of the (logarithmic) integrand in (25) superimposed on the contours of the Gaussian approximation underlying (29). Any deviation between the two sets of contours is only visible near the corners, i.e., for \mathbf{p}_2 far away from the estimate $\hat{\mathbf{p}}_{2|2}^\theta$.

the calculated covariance [16]. This effect can have a huge detrimental impact on evaluation of the hypothesis probabilities $\Pr(\theta | \mathbf{Z}_{1:2})$, since these essentially are exponential in the underlying Mahalanobis distances. Any solution to this problem must either inflate the covariance to correspond to the actual estimation error, or it must improve the estimation accuracy so that it actually reaches the expected covariance. We have found that the NG technique is quite successful in fulfilling the second option.

A. EKF-based pose estimation

In this stage, we linearize the hypothesis-conditioned estimation problem around the predicted state vector $\xi_{1|1}^\theta$. The linearization is carried out by means of the Jacobian

$$\mathbf{H}^\theta = \begin{bmatrix} \mathbf{H}_p^{\theta(1)} & \mathbf{H}_x^{\theta(1)} & & \\ \vdots & & \ddots & \\ \mathbf{H}_p^{\theta(n)} & & & \mathbf{H}_x^{\theta(n)} \end{bmatrix} \Big|_{\xi=\xi_{1|1}^\theta} \quad (38)$$

with sub-matrices

$$\begin{aligned} \mathbf{H}_p^{\theta(i)} &= \mathcal{D}_{\mathbf{p}_2} \mathbf{h}(\mathbf{y}_1^{\theta_1(i)}, \mathbf{p}_2) = \mathcal{D}_g \mathbf{f} \mathcal{D}_{\mathbf{p}_2} \mathbf{g} \\ \mathbf{H}_x^{\theta(i)} &= \mathcal{D}_{\mathbf{y}_1^{\theta_1(i)}} \mathbf{h}(\mathbf{y}_1^{\theta_1(i)}, \mathbf{p}_2) = \mathcal{D}_g \mathbf{f} \mathcal{D}_{\mathbf{y}_1^{\theta_1(i)}} \mathbf{g} \end{aligned} \quad (39)$$

where

$$\begin{aligned} \mathcal{D}_g \mathbf{f} &= \begin{bmatrix} x/\rho & y/\rho \\ -y/\rho^2 & x/\rho^2 \end{bmatrix} \\ \mathcal{D}_{\mathbf{p}_2} \mathbf{g} &= \begin{bmatrix} -R(-\psi_2) & S(\psi_2)(\mathbf{y}_1^{\theta_1(i)} - \boldsymbol{\rho}_2) \end{bmatrix} \\ \mathcal{D}_{\mathbf{y}_1^{\theta_1(i)}} \mathbf{g} &= R(-\psi_2). \end{aligned} \quad (40)$$

Here ρ , x and y are given according to $\rho = \|\mathbf{y}_1^{\theta_1(i)} - \boldsymbol{\rho}_2\|_2$ and $[x, y]^T = R(-\psi_2)(\mathbf{y}_1^{\theta_1(i)} - \boldsymbol{\rho}_2)$. Furthermore, we define the combined measurement mapping

$$\tilde{\mathbf{h}}(\xi^\theta) = [\mathbf{h}(\mathbf{x}^1, \mathbf{p}_2)^T, \dots, \mathbf{h}(\mathbf{x}_1^n, \mathbf{p}_2)^T]^T \quad (41)$$

and the corresponding covariance

$$\mathbf{R}^\theta = \mathbf{I}_n \otimes \mathbf{R}. \quad (42)$$

An EKF with output $\xi_{2|2}^\theta$ and $\mathbf{P}_{2|2}^\theta$ can then be constructed as follows:

$$\mathbf{W}^\theta = \mathbf{P}_{1|1}^\theta (\mathbf{H}^\theta)^T (\mathbf{R}^\theta + \mathbf{H}^\theta \mathbf{P}_{1|1}^\theta (\mathbf{H}^\theta)^T)^{-1} \quad (43)$$

$$\boldsymbol{\nu}^\theta = \mathbf{z}_2^\theta - \tilde{\mathbf{h}}(\xi_{1|1}^\theta) \quad (44)$$

$$\xi_{2|2}^\theta = \xi_{1|1}^\theta + \mathbf{W}^\theta \boldsymbol{\nu}^\theta \quad (45)$$

$$\mathbf{P}_{2|2}^\theta = (\mathbf{I} - \mathbf{W}^\theta \mathbf{H}^\theta) \mathbf{P}_{1|1}^\theta. \quad (46)$$

We can partition the EKF-estimate into pose and landmark components according to $\xi_{2|2}^\theta = [(\mathbf{p}_{2|2}^\theta)^T, (\mathbf{x}_{2|2}^\theta)^T]^T$, and the corresponding covariance can be partitioned as

$$\mathbf{P}_{2|2}^\theta = \begin{bmatrix} \mathbf{P}_{\mathbf{p}}^\theta & \mathbf{P}_{\mathbf{p}\mathbf{x}}^\theta \\ (\mathbf{P}_{\mathbf{p}\mathbf{x}}^\theta)^T & \mathbf{P}_{\mathbf{x}}^\theta \end{bmatrix}.$$

B. Natural gradient optimization

The NG is an optimization technique similar to the more well-known optimization method of Newton. Instead of the Hessian of the cost function, the NG uses a Riemannian metric tensor which is “naturally defined from the characteristics of the parameter space” [9]. By using the information matrix $\mathbf{J}^\theta(\mathbf{p}_{2|2}^\theta)$ as the Riemannian tensor, one can derive the single-iteration optimization scheme:

$$\hat{\mathbf{p}}_{2|2}^\theta = \mathbf{p}_{2|2}^\theta - (\mathbf{J}^\theta(\mathbf{p}_{2|2}^\theta))^{-1} \nabla \quad (47)$$

where

$$\begin{aligned} \nabla &= (\mathbf{p}_2 - \boldsymbol{\mu}_q)^T \mathbf{P}^{-1} \\ &+ \frac{1}{2} \sum_{i \in \theta} \left(\mathbf{d}_1 + \mathbf{d}_2 + \text{vec}((\boldsymbol{\Sigma}^{\theta(i)})^{-1})^T \mathcal{D}_{\mathbf{p}_2} \boldsymbol{\Sigma}^{\theta(i)} \right) \\ \mathbf{d}_1 &= ((\boldsymbol{\Sigma}^{\theta(i)})^{-1} \boldsymbol{\nu}^{\theta(i)})^T \mathcal{D}_{\mathbf{p}_2} \boldsymbol{\nu}^{\theta(i)} \\ \mathbf{d}_2 &= (\boldsymbol{\nu}^{\theta(i)})^T [((\boldsymbol{\nu}^{\theta(i)})^T \otimes \mathbf{I}_2) \mathcal{D}_{\mathbf{p}_2} [(\boldsymbol{\Sigma}^{\theta(i)})^{-1}] \\ &\quad + (\boldsymbol{\Sigma}^{\theta(i)})^{-1} \mathcal{D}_{\mathbf{p}_2} \boldsymbol{\nu}^{\theta(i)}] \\ \mathcal{D}_{\mathbf{p}_2} (\boldsymbol{\Sigma}^{\theta(i)})^{-1} &= -((\boldsymbol{\Sigma}^{\theta(i)})^{-1} \otimes (\boldsymbol{\Sigma}^{\theta(i)})^{-1}) \mathcal{D}_{\mathbf{p}_2} \boldsymbol{\Sigma}^{\theta(i)} \\ \mathcal{D}_{\mathbf{p}_2} \boldsymbol{\Sigma}^{\theta(i)} &= [\mathbf{0}_{4 \times 2}, \text{vec}(\mathbf{U}^{\theta(i)})]. \end{aligned}$$

The landmark estimates and the covariance are not optimized during this stage, and these quantities therefore remain as in (45) and (46).

Figure 3 shows the normalized estimation error $(\hat{\mathbf{p}}_{2|2} - \mathbf{p}_2^{\text{true}})^T \mathbf{P}_{\mathbf{p}}^{-1} (\hat{\mathbf{p}}_{2|2} - \mathbf{p}_2^{\text{true}})$ averaged over 200 Monte-Carlo runs for a scenario with known data association. One can see that the EKF is increasingly inconsistent for finer sensor resolution, while the NG achieves consistency for all sensor resolutions.

VI. HYPOTHESIS SEARCH BY CLIQUE DETECTION

In practical MHT implementations, the key challenge is to search for promising hypotheses without brute force enumeration. This search is typically carried out by assignment methods based on integer programming [1] or the auction algorithm [17]. These methods exploit a key feature of multiple hypothesis *tracking*, namely that the cost of a hypothesis can be written as a sum of terms contributed by the measurements involved. In multi-hypothesis *SLAM* this is no longer possible, since the correspondences become co-dependent when the

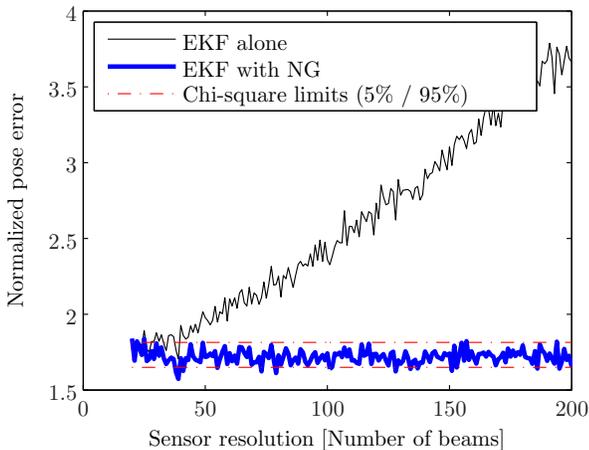


Fig. 3: Consistency of NG optimization.

uncertainty of \mathbf{p}_2 is taken into account. Also, assignment methods such as integer programming or the auction algorithm only aim to discover a single best hypothesis, or possibly the N best hypotheses. They do not make any attempt at exploring the hypothesis space beyond such limitations.

In this paper we solve the search problem through a 3-stage procedure consisting of validation gating, clique-detection and hypothesis expansion.

A. Validation gating

First of all, validation gating similar to the individual compatibility test of [7] is used in order to keep the number of correspondences on a manageable level. More precisely, \mathbf{z}_2^j may possibly be associated with \mathbf{z}_1^i only if

$$(\mathbf{z}_2^j - \mathbf{h}(\mathbf{z}_1^i, \boldsymbol{\mu}_q))^T (\mathbf{S}^i)^{-1} (\mathbf{z}_2^j - \mathbf{h}(\mathbf{z}_1^i, \boldsymbol{\mu}_q)) < g^2 \quad (48)$$

where g is the number of standard deviations tolerated and

$$\mathbf{S}^i = \mathbf{H}_p^i \mathbf{P} (\mathbf{H}_p^i)^T + \mathbf{H}_x^i \mathbf{Y}_1^i (\mathbf{H}_x^i)^T.$$

The Jacobians \mathbf{H}_p^i and \mathbf{H}_x^i are found according to (39). The output of the gating procedure can be represented by a matrix $\mathbf{G} \in \{0, 1\}^{m_1 \times m_2}$ whose entry number (i, j) is one if \mathbf{z}_1^i may be associated with \mathbf{z}_2^j , and zero otherwise.

B. Clique detection

The search problem can be addressed by treating the individual correspondences as nodes in a graph, whose edges represent distance between the correspondences. For the FBSM problem discussed in this paper, a correspondence consists of two measurements. This yields one free parameter during displacement estimation. The locus of all possible \mathbf{p}_2 for a given correspondence can be visualized as a helix. Following this line of thought, the nodes of the correspondence graph correspond to helices, and the edges correspond to MSS's. The edge weights are related to the minimal distances between helices. One would expect that all the helices involved in a good hypothesis should intersect a limited volume near the correct displacement vector $\mathbf{p}_2^{\text{true}}$.

Working with helices is problematic for two reasons. First, the construction of helices is fundamentally a non-probabilistic

method. Second, inter-helix distances would have to be evaluated using a Riemannian metric, but it is not immediately clear how this one should be defined or minimized.

Instead, we establish the edges of the correspondence graph using normalized re-projection error. For any MSS θ which can be generated from \mathbf{G} we find its normalized re-projection error as

$$e(\theta) = (\boldsymbol{\nu}^\theta)^T (\mathbf{H}^\theta \mathbf{P}_{2|2}^\theta (\mathbf{H}^\theta)^T + \mathbf{R}^\theta)^{-1} \boldsymbol{\nu}^\theta \quad (49)$$

Only those edges whose normalized re-projection error is below some threshold t are included in the graph. Furthermore, if the number of edges exceeds some constant n , then we only retain the n best edges.

In order to find the maximal cliques of this graph we use the Bron-Kerbosch method [10] as implemented in [18].

C. Hypothesis expansion

The collection of *maximal* cliques does not necessarily contain the true hypothesis θ^{true} . One can nevertheless rest assured that θ^{true} will be a sub-clique of one of the maximal cliques, insofar as t and n are large enough. Therefore, for each maximal clique θ , we also include all sub-cliques which contain $n(\theta) - d$ or more nodes in the hypothesis collection. We refer to d as the expansion depth.

VII. TEST DESIGN AND SIMULATION RESULTS

A. Simulation setup

We investigate the performance of CDSM through Monte-Carlo simulations which are designed to mimic output from a multi-beam sonar with maximum range 60m, and total azimuth coverage of 120° . The vehicle's surge velocity is uniformly drawn from the interval $[0 \text{ knots}, 5 \text{ knots}]$, while the time interval between the two scans is uniformly drawn from the interval $[1s, 3s]$, so as to mimic a realistic AUV application.

Landmarks are placed at random around the vehicle, and detected with various detection probabilities in $[0, 1]$, independently between the scans. For any detected landmark, the measurement noise is given by $\mathbf{R} = \text{diag}([0.125, 5.7 \cdot 10^{-4}])$, corresponding to a sensor resolution of 60×31 cells [12]. Clutter measurements are drawn with a spatial distribution being uniform over the FOV in polar coordinates. The number of clutter measurements are drawn according to Poisson distributions corresponding to various false alarm rates in $[10^{-10}, 10^{-1}]$.

Having generated several hundreds of thousands of such scenarios, we pick the first 200 Monte-Carlo runs which satisfy various constraints given by the number n of landmark measurements, by the average number $\bar{\phi} = (\phi_1 + \phi_2)/2$ of clutter measurements, and by the relative pose angle ψ_2 between the two scans. We investigate a total of 16 such scenarios (see Tables I - IV).

The pose prior is given by $\boldsymbol{\mu}_q = \mathbf{0}_{3 \times 1}$ and $\mathbf{P} = \text{diag}([(5 \text{ m})^2, (2 \text{ m})^2, (30^\circ)^2])$. Normalized re-projection error $e(\theta)$ is thresholded at $t = 4$ standard deviations, and validation gating is done with $g = 3$. Maximally $n = 200$ correspondences are included in the correspondence graph.

Notice that CDSM makes minimal assumptions regarding false-alarm rate or detection probability. Thus the model assumed by CDSM differs from the simulation model with regard to these quantities.

B. Benchmark methods

We compare CDSM with 5 other methods for scan-matching and data association in SLAM.

1) *JCBB*: We use a Matlab implementation of JCBB which was downloaded from www.robots.ox.ac.uk/~SSS06. It should be noted that the original implementation did not give satisfactory performance for large $|\psi_2|$, due to usage of converted measurements \mathbf{y}_2^i when computing joint compatibility and Mahalanobis distances. The program is therefore rewritten so that the polar measurements \mathbf{z}_2^i are used instead of the converted measurements \mathbf{y}_2^i in these tasks. JCBB's estimate of \mathbf{p}_2 is found using the same technique as CDSM uses.

2) *RANSAC*: A generic RANSAC method works by drawing random MSS's until k iterations have been executed. During each iteration, all correspondences are tested for compatibility with the current MSS, and added to the current hypothesis if deemed compatible. A score function is then evaluated to test whether the new hypothesis is better than the previously best hypothesis. The number k is dynamically updated during each iteration, so as to reflect how many more iterations are deemed necessary to find the true solution with a given probability p ($p = 0.999$ is used in our simulations):

$$k = \frac{1 - p}{1 - (\text{card}(\theta^*)/M)^s} \quad (50)$$

Here M is the total number of feasible correspondences, $s = 2$ is the size of an MSS, and $\text{card}(\theta^*)$ is the cardinality of the best hypothesis θ^* found so far.

RANSAC requires us to specify an estimator of \mathbf{p}_2 , an error function for whether a correspondence fits the MSS, and a hypothesis score function. In "standard" RANSAC, we use the formulas suggested by [19] to estimate \mathbf{p}_2 . As the error function, we use re-projection error, defined as

$$\text{error}(\theta) = \frac{1}{n} \sum_{i=1}^n \|\mathbf{R}(-\hat{\psi})\mathbf{z}_1^{\theta_1(i)} - \hat{\rho} - \mathbf{z}_2^{\theta_2(i)}\|_2 \quad (51)$$

When testing a correspondence's fit with an MSS, this function is evaluated for each the resulting 3 pairs of correspondences. The correspondence is added to the hypothesis if the average of all 3 reprojection errors is smaller than a threshold $T = 2\text{m}$. As hypothesis score we use cardinality of hypotheses. Whenever a tie is encountered, the hypothesis set with the lowest re-projection error is chosen.

3) *Probabilistic RANSAC (PSAC)*: As an alternative to the "standard" RANSAC method we also suggest a "probabilistic" version of RANSAC which utilizes the machinery of Section IV. In this method, \mathbf{p}_2 is found according to the KF-based formulas of Section V. As error function, we use normalized reprojection error as defined in (49). Hypothesis cardinality is used as the primary score function, but whenever a tie is encountered, we use the posterior hypothesis score of Section IV.

4) *pIC*: Standard scan-matching methods work in terms of a two-step iterative process which is repeated till convergence. In step 1, a measure of the plausibility of each tentative correspondence is calculated. In step 2, the displacement between the two scans is calculated according to a least-squares criterion which puts most emphasis on the most plausible correspondences. The most popular methods of this kind are the iterative closest point (ICP) method [20, 21] and the probabilistic iterative correspondence (pIC) method [22, 23]. The latter is specifically tailored towards working with measurements received by range-bearing sensors. In our simulations we use pIC as described in [22] with maximally 8 iterations.

C. Performance measures

The output of our simulations is analyzed in terms of several performance measures, of which two are reported here.

First, we investigate in terms of hard thresholding whether the scan-matching succeeds or not. For CDSM, JCBB, RANSAC and PSAC, a match is declared successful if the top non-empty hypothesis has at least $\max(2, n^{\text{true}}/2)$ correct correspondences, and its pose estimate $\hat{\mathbf{p}}_{2|2}^\theta$ is within 3 standard deviations of the true pose displacement $\mathbf{p}_2^{\text{true}}$. That is, we require that

$$(\hat{\mathbf{p}}_{2|2}^\theta - \mathbf{p}_2^{\text{true}})^\top \mathbf{P}_{\mathbf{p}, \text{true}}^{-1} (\hat{\mathbf{p}}_{2|2}^\theta - \mathbf{p}_2^{\text{true}}) < 9 \quad (52)$$

where $\mathbf{P}_{\mathbf{p}, \text{true}}$ is the posterior pose covariance of the *true hypothesis*. For pIC we only require that its final estimate must satisfy (52).

Second, we investigate the consistency properties of CDSM and JCBB. The concept of consistency has several different meanings in estimation theory. In this paper, we understand the concept as it is used in [14]: That the estimation error should have magnitude commensurate with the corresponding covariance that is yielded by the estimator. In other words, a consistent estimator has the right degree of optimism. This concept of consistency is fundamentally moment-based, and is consequently most directly applicable to Gaussian (or at least unimodal) estimation problems. For more general (possibly multimodal) estimation problems, we propose the following definition of consistency in terms of the pdf:

Let E be a solution to a Bayesian estimation problem with data \mathbf{z} and with state $\mathbf{x} \in \mathbb{R}^n$. Assume that the output of E can be represented in terms of a pdf $f(\mathbf{x}|\mathbf{z})$. For any realization of \mathbf{x} and \mathbf{z} , let $f_{\text{max}} = \max_{\mathbf{x}'} f(\mathbf{x}'|\mathbf{z})$ and let $f_{\text{true}} = f(\mathbf{x}|\mathbf{z})$. In other words, f_{true} is f evaluated at the true \mathbf{x} . Also, let u be a Gaussian random variable with distribution $\mathcal{N}(0, 1)$. Then we say that E is *pdf-consistent* if the distribution of $f_{\text{true}}/f_{\text{max}}$ is commensurate with the distribution of the ratio $\mathcal{N}(u; 0, 1)/\mathcal{N}(0; 0, 1)$.

If E is the full solution to the underlying estimation problem, then the obvious candidate for f would be the posterior pdf. If E is, say, an EKF, then it would be most reasonable to let f be the Gaussian whose moments are given by the EKF. For CDSM, f would be the posterior pdf

$$f(\mathbf{p}_2 | \mathbf{Z}_{1:2}) = \sum_{\theta} \text{Pr}(\theta | \mathbf{Z}_{1:2}) \int f(\boldsymbol{\xi}^\theta | \theta, \mathbf{Z}_{1:2}) d\mathbf{x}^\theta. \quad (53)$$

For JCBB, we represent its chosen hypothesis θ by the Gaussian $\mathcal{N}(\mathbf{p}_2; \hat{\mathbf{p}}_{2|2}^\theta, \mathbf{P}_{\mathbf{p}}^\theta)$.

The definition of pdf-consistency is somewhat vague. The point is that one should investigate the general pdf of the pdf-ratios, and particularly the lower tail of this pdf, in order to make an assessment of pdf-consistency. This can for example be done by means of a box-plot.

D. Simulation results

Success-rates for increasingly difficult scenarios are displayed in Tables I - IV. We make the following four comments on these results:

The success rates of CDSM and JCBB are very similar for practically all scenarios. They are generally quite high, but not perfect. Even in the simplest case (Table I) there are several Monte-Carlo runs where both CDSM and JCBB fail to choose the correct hypothesis. However, in the vast majority of these cases, a hypothesis which satisfies our success criteria can be found among the top 10 CDSM hypotheses. Such hedging is not provided by JCBB since it only returns one hypothesis.

Both CDSM and JCBB perform reasonably well for up to three times as much clutter as landmarks. For higher clutter rates, performance deteriorates rapidly. An interesting, but highly non-trivial topic for future research may be to develop rigorous Cramer-Rao lower bounds based on model parameters such as sensor resolution and clutter rate.

The performance of both RANSAC methods (RANSAC and PSAC) is variable. It is possible that PSAC could be tuned to yield as good success rates as CDSM/JCBB, but this would require many bells and whistles. As for standard RANSAC, we notice that this is the only method which does not use an EKF-based pose estimation technique. Consequently, standard RANSAC suffers from a significant deterioration as $|\psi_2|$ increases. A similar deterioration could also be observed for JCBB when converted measurements were used to compute joint compatibility and Mahalanobis distances.

In most cases, pIC fails to provide adequate estimates. This could possibly be because pIC depends on continuous features such as walls, or because there are too few landmarks in the simulated scenarios. However, even when pIC violates the success criterion, it is not necessarily entirely lost. Often pIC yields near-acceptable estimates, but local optima prevent the method from converging to sufficiently accurate estimates.

In order to investigate consistency, Figure 4 displays box-plots of pdf-ratios for the scenarios corresponding to Table II (CDSM 1 and JCBB 1) and Table III (CDSM 2 and JCBB 2). For the easy scenario, both CDSM and JCBB show a similar and quite acceptable overall performance, although JCBB has many more outliers than CDSM. For the difficult scenario, the consistency properties of CDSM are not considerably changed, while the consistency of JCBB deteriorates significantly. More concretely, we can from Figure 4 conclude that CDSM always considers the true pose to be at least 10^{-4} times as likely as its preferred pose estimate. On average it is considered to be 0.25 times as likely as its preferred pose estimate. For JCBB, the corresponding numbers are 10^{-317} (far beyond the lower limit of the figure) and 0.1. In practical terms, this means that there will be many occasions where JCBB does not give the correct pose vector any serious consideration, and this is of course problematic from a robustness perspective.

TABLE I: Successes for $5 \leq n < 10$, $\bar{\phi} < 10$

	CDSM	JCBB	RANSAC	PSAC	pIC
$ \psi_2 < 2^\circ$	94.0%	93.5%	90.5%	74.5%	31.0%
$ \psi_2 \in [2^\circ, 8^\circ)$	92.5%	95.5%	55.5%	73.5%	30.5%
$ \psi_2 \in [8, 32^\circ)$	94.5%	96.0%	0.0%	81.5%	19.5%
$ \psi_2 \geq 32^\circ$	95.0%	95.0%	0.0%	83.5%	7.0%

TABLE II: Successes for $n < 5$, $\bar{\phi} < 10$

	CDSM	JCBB	RANSAC	PSAC	pIC
$ \psi_2 < 2^\circ$	92.0%	90.0%	86.5%	66.5%	30.5%
$ \psi_2 \in [2^\circ, 8^\circ)$	86.0%	87.0%	64.5%	65.0%	29.0%
$ \psi_2 \in [8, 32^\circ)$	93.5%	91.5%	18.0%	74.0%	19.0%
$ \psi_2 \geq 32^\circ$	88.0%	88.0%	8.5%	71.0%	8.5%

TABLE III: Successes for $n < 5$, $10 \leq \bar{\phi} < 15$

	CDSM	JCBB	RANSAC	PSAC	pIC
$ \psi_2 < 2^\circ$	81.5%	83.0%	71.0%	35.5%	12.5%
$ \psi_2 \in [2^\circ, 8^\circ)$	82.0%	73.5%	43.5%	31.5%	10.5%
$ \psi_2 \in [8, 32^\circ)$	79.5%	73.0%	4.5%	33.5%	7.0%
$ \psi_2 \geq 32^\circ$	71.5%	72.5%	0.0%	45.0%	0.5%

TABLE IV: Successes for $n < 5$, $15 \leq \bar{\phi} < 20$

	CDSM	JCBB	RANSAC	PSAC	pIC
$ \psi_2 < 2^\circ$	52.0%	41.0%	50.0%	11.5%	9.0%
$ \psi_2 \in [2^\circ, 8^\circ)$	52.5%	52.0%	20.5%	9.5%	4.0%
$ \psi_2 \in [8, 32^\circ)$	47.5%	46.5%	1.0%	12.0%	2.5%
$ \psi_2 \geq 32^\circ$	25.0%	23.0%	0.0%	8.0%	0.0%

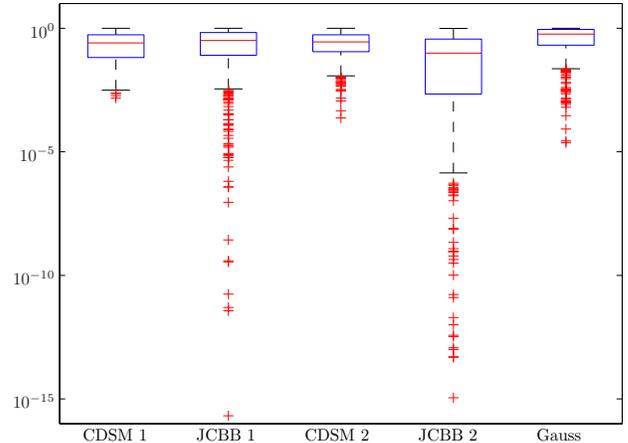


Fig. 4: Consistency of CDSM versus JCBB.

TABLE V: Frame displacements

		Frame 1	Frame 2	Frame 3	Frame 4	Frame 5
Frame 1	Δx	0.0 m	2.8 m	5.3 m	5.3 m	9.3 m
	Δy	0.0 m	-6.3 m	-12.0 m	12.0 m	-20.8 m
	$\Delta \psi$	0.0°	-1.0°	-20.7°	-11.8°	-44.8°
Frame 2	Δx		0.0 m	2.6 m	2.6 m	6.7 m
	Δy		0.0 m	-5.6 m	-5.6 m	-14.4 m
	$\Delta \psi$		0.0°	-19.7°	-10.8°	-43.8°
Frame 3	Δx			0.0 m	0.0 m	6.8 m
	Δy			0.0 m	0.0 m	-6.9 m
	$\Delta \psi$			0.0°	8.9°	-24.1°
Frame 4	Δx				0.0 m	5.7 m
	Δy				0.0 m	-7.9 m
	$\Delta \psi$				0.0°	-33.0°
Frame 5	Δx					0.0 m
	Δy					0.0 m
	$\Delta \psi$					0.0°

VIII. IMPLEMENTATION ON REAL SONAR DATA

In order to investigate real world applicability, CDSM is also implemented and compared to JCBB on real sonar data recorded by a 650 kHz Micron DST forward looking sonar (FLS) at the Republic Of Singapore Yacht Club (RSYC). In this experiment, the sonar head is placed at 5 different poses in a marina environment. Several data scans are recorded at each pose. The sonar can in this environment observe backscatter from several solid poles, as well as from boats, the seafloor and other miscellaneous surroundings. Successful scan-matching requires CDSM/JCBB to correctly match the poles observed in one scan with the poles observed in another scan.

A. Measurement extraction

For the feature-based approach to be applicable, point features must be extracted from the sonar scans. This is done through a two-step procedure comprising detection and clustering. The latter step is included in order to ensure the validity of the assumption that at most one measurement originates from each landmark.

For detection, we use a simple heuristic scheme: For each bearing (a.k.a. beam, scan-line) the mean and standard deviation for the amplitudes of the outermost 25 cells are calculated. From this, a threshold T_{outer} is calculated by inverting the corresponding Gaussian cumulative distribution function with design false alarm rate 0.0005. For the innermost resolution cell, the threshold T_{inner} is set to 5 times T_{outer} , and for resolution cells number 1 – 40 the range-dependent threshold is found by linear interpolation between T_{outer} and T_{inner} . This scheme is motivated by two observations: First, the amplitude value which is exceeded by 1% of the resolution cells appears to be a negative linear function of range for $r < 40$ m, and a constant function of range for $r > 40$ m. Second, the “average” background increases from near zero for small ranges to a near-Gaussian steady-state component for $r > 30$ m.

For clustering we use a scheme based on the clustering method that was proposed in [12]. Briefly speaking, the method of [12] attempts to bring together adjacent detected cells into “blobs”. Clustering was in [12] done in 3 stages: Clustering of cells along the bearing direction, partitioning of bearing-clusters into smaller clusters when deemed appropriate, and clustering of bearing-clusters along the range direction. These 3 stages are carefully explained in [12], and the details are

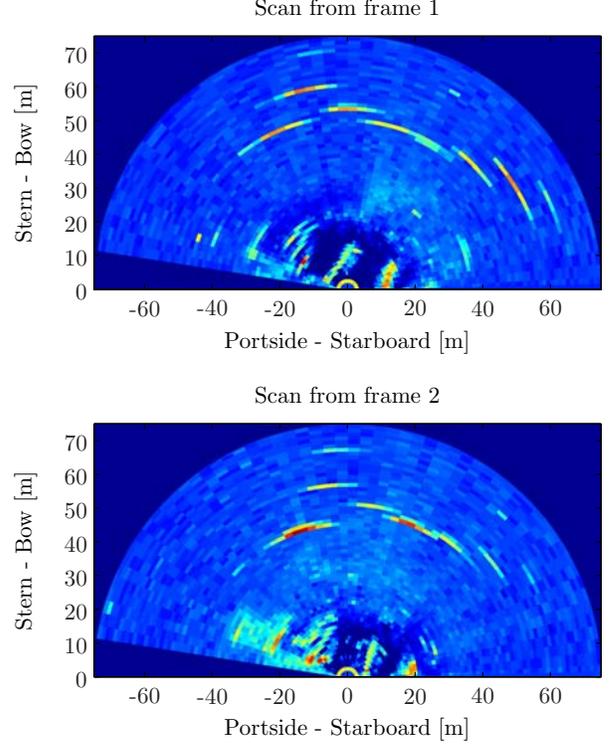


Fig. 5: Full data scans recorded with the sonar in different poses. Backscatter from the poles in the RSYC marina can be observed at 40 – 60 m range. A slight displacement between the scans can easily be observed, and the objective of scan-matching is to estimate this displacement.

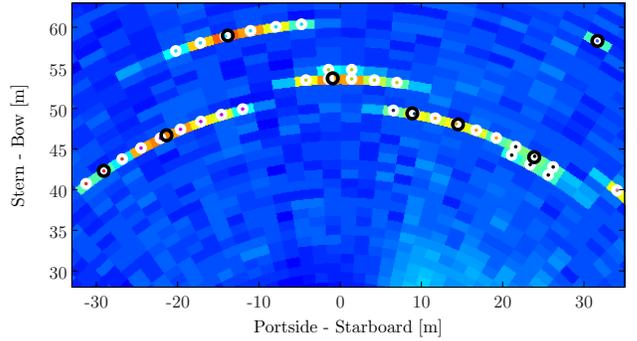


Fig. 6: Illustration of the measurement extraction process. 35 detected pixels can be connected into 5 connected clusters, which give rise 8 extracted point measurements.

therefore not repeated here. In addition to these 3 stages we include a 4th stage: Partitioning of range-clusters into smaller clusters when deemed appropriate. This is done whenever a range-cluster contains two or more non-adjacent peaks. Each cell is then assigned to its nearest peak cell. Finally, centroids are obtained by a weighted average over all cells in each cluster. Figure 6 illustrates how the clustering process reduces the number of detected measurements considerably.

TABLE VI: Success rates

		Frame 1	Frame 2	Frame 3	Frame 4	Frame 5
Frame 1	CDSM	87.9%	92.5%	75.2%	71.9%	0.5%
	JCBB	86.7%	91.3%	79.2%	70.3%	1.1%
	CDSM*	100.0%	100.0%	100.0%	97.4%	14.0%
Frame 2	CDSM		100.0%	97.7%	86.2%	58.4%
	JCBB		99.6%	96.9%	86.2%	49.9%
	CDSM*		100.0%	100.0%	98.4%	97.9%
Frame 3	CDSM			100.0%	83.6%	89.9%
	JCBB			99.8%	83.1%	90.6%
	CDSM*			100.0%	97.4%	99.9%
Frame 4	CDSM				72.7%	56.7%
	JCBB				72.7%	56.5%
	CDSM*				97.4%	86.7%
Frame 5	CDSM					94.7%
	JCBB					94.9%
	CDSM*					100.0%

B. Performance measures and results

Also for the real data we use success rates to measure performance. We report 3 success rates: Whether the top non-empty CDSM hypothesis is deemed good enough, whether JCBB's hypothesis is deemed good enough, and whether *any* hypothesis in CDSM's hypothesis collection is deemed good enough, referred to as CDSM* in Table VI. The success criterion is defined slightly differently here: We require at least 3 landmarks to have been correctly identified, and we require the deviation from ground truth to be less than 1 m, 1 m and 3° in x , y and ψ , respectively. Also, due to the relatively large lateral displacements (see Table V) we have increased the prior covariance to $\mathbf{P} = \text{diag}([(14\text{m})^2, (14\text{m})^2, (30^\circ)^2])$.

In Table VI we report success rates for the various frame displacements. Again, the success-rates of CDSM and JCBB are strikingly similar, and again the hedging capability of CDSM is evident. Only when attempting to match Frame 1 and Frame 5 does CDSM entirely fail. Both JCBB and CDSM yield poor results for the matching of Frame 2 and Frame 5 as well. This gives some idea of the limitations of the proposed approach: The underlying linearizations become inadequate for lateral displacements exceeding 15 m and displacement angles exceeding 40° .

IX. CONCLUSION

In this paper we have developed an MHT-based solution to the 2-frame SLAM problem. To the best of our knowledge, this is the first application of the MHT framework to autonomous navigation reported in the literature. The proposed solution performs roughly as well as JCBB, but has better robustness and consistency properties.

There is a vast potential for future research based on this paper. It would be desirable to extend our work to general SLAM with > 2 data frames. Research on the tracking of moving objects using moving sensors may also possibly make use of ideas and concepts from this paper. When restricted to two frames, extensions of our work could address the so-called kidnapped robot problem. The hypothesis search technique used in this paper could possibly be improved. Finally, on the theoretical side, it is desirable to investigate exactly how the proposed approach relates to FISST, and to which extent it can be said to represent an optimal solution.

ACKNOWLEDGMENT

The authors would like to thank Mr. Chew Jee Loong for recording the sonar data studied in Section VIII.

REFERENCES

- [1] W. S. Wijesoma, L. D. L. Perera, and M. D. Adams, "Toward multidimensional assignment data association in robot localization and mapping," *IEEE Transactions on Robotics*, vol. 22, no. 2, pp. 350–365, Apr. 2006.
- [2] S. J. Davey, "Simultaneous localization and map building using the probabilistic multi-hypothesis tracker," *IEEE Transactions on Robotics*, vol. 23, no. 2, pp. 271–280, Apr. 2007.
- [3] M. Montemerlo, "FastSLAM: A factored solution to the simultaneous localization and mapping problem with unknown data association," Ph.D. dissertation, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, June 2003.
- [4] C. S. Lee, D. Clark, and J. Salvi, "SLAM with single cluster PHD filters," in *Proceedings of 2012 IEEE ICRA*, St. Paul, MN, USA, May 2012, pp. 2096–2101.
- [5] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, Dec. 1979.
- [6] E. Brekke and M. Chitre, "Relationship between finite set statistics and the multiple hypothesis tracker," December 2012, submitted to *IEEE Transactions on Signal Processing*.
- [7] J. Neira and J. D. Tardós, "Data association in stochastic mapping using the joint compatibility test," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 6, pp. 890–897, Dec. 2001.
- [8] U. Frese, "Interview: Is SLAM solved?" *Künstliche Intelligenz*, vol. 24, pp. 255–257, 2010.
- [9] S. Amari and S. C. Douglas, "Why natural gradient?" in *Proceedings of 1998 IEEE ICASP*, Seattle, WA, USA, May 1998, pp. 1213–16.
- [10] C. Bron and J. Kerbosch, "Finding all cliques of an undirected graph," *Communications of the ACM*, vol. 16, no. 9, pp. 575–7, Sep. 1973.
- [11] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," *Autonomous robot vehicles*, pp. 167–193, 1990.
- [12] E. Brekke, O. Hallingstad, and J. Glattetre, "The signal-to-noise ratio of human divers," in *Proc. OCEANS'10*, Sydney, Australia, May 2010.
- [13] S. Mori, C.-Y. Chong, E. Tse, and R. Wishner, "Tracking and classifying multiple targets without a priori identification," *IEEE Transactions on Automatic Control*, vol. 31, no. 5, pp. 401–409, 1986.
- [14] Y. Bar-Shalom and X. R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. Storrs, CT, USA: YBS Publishing, 1995.
- [15] B. Porat and B. Friedlander, "Computation of the exact information matrix of gaussian time series with stationary random components," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 1, pp. 118–30, Feb. 1986.
- [16] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, "Observability-based rules for designing consistent EKF SLAM estimators," *The International Journal of Robotics Research*, vol. 29, no. 5, pp. 502–528, Apr. 2010.
- [17] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.
- [18] J. Wildman, "Bron-Kerbosch maximal clique finding algorithm," Matlab Central File Exchange. Retrieved 27 October 2011.
- [19] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America A*, vol. 4, no. 4, pp. 629–42, Apr. 1987.
- [20] P. J. Besl and H. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–56, Feb. 1992.
- [21] J. Nieto, T. Bailey, and E. Nebot, "Recursive scan-matching SLAM," *Robotics and Autonomous Systems*, vol. 55, no. 1, pp. 39–49, Jan. 2007.
- [22] L. Montesano, J. Minguez, and L. Montano, "Probabilistic scan matching for motion estimation in unstructured environments," in *Proceedings of IROS 2005*, Edmonton, AB, Canada, Aug. 2005, pp. 3499–504.
- [23] E. Hernandez, P. Ridao, D. Ribas, and J. Batlle, "MSISpIC: A probabilistic scan matching algorithm using a mechanical scanned imaging sonar," *Journal of Physical Agents*, vol. 3, no. 1, pp. 3–11, Jan. 2009.