

# Efficient Constellation-Based Map-Merging for Semantic SLAM

Kristoffer M. Frey<sup>1</sup>, Ted J. Steiner<sup>2</sup>, and Jonathan P. How<sup>1</sup>

**Abstract**—Data association in SLAM is fundamentally challenging, and handling ambiguity well is crucial to achieve robust operation in real-world environments. When ambiguous measurements arise, conservatism often mandates that the measurement is discarded or a new landmark is initialized rather than risking an incorrect association. To address the inevitable “duplicate” landmarks that arise, we present an efficient map-merging framework to detect duplicate *constellations* of landmarks, providing a high-confidence loop-closure mechanism well-suited for object-level SLAM. This approach uses an incrementally-computable approximation of landmark uncertainty that only depends on local information in the SLAM graph, avoiding expensive recovery of the full system covariance matrix. This enables a search based on geometric consistency (GC) (rather than full joint compatibility (JC)) that inexpensively reduces the search space to a handful of “best” hypotheses. Furthermore, we reformulate the commonly-used interpretation tree to allow for more efficient integration of clique-based pairwise compatibility, accelerating the branch-and-bound max-cardinality search. Our method is demonstrated to match the performance of full JC methods at significantly-reduced computational cost, facilitating robust object-based loop-closure over large SLAM problems.

## I. INTRODUCTION

The rise of single-shot object detectors [1]–[3] has led to interest in extensions of classic SLAM algorithms to include semantically-meaningful landmarks. For mobile robots operating in the real world, the ability to detect and localize semantic objects such as cars or street signs in the environment is vital for safe and effective behavior. Besides the benefits for motion planning, the inclusion of these semantic landmarks in the SLAM factor-graph also provides opportunities to improve data association and loop-closure, fundamental challenges in SLAM [4]. Compared to generic point features, semantic landmarks represent whole “objects” in the world, making them highly distinctive. Furthermore, modern object detectors [1], [2] are more robust to viewpoint and lighting variation than generic image-space descriptors such as SURF [5]. Loop-closure is crucial for autonomous systems operating without the aid of GPS or other localization infrastructure, as any pure-odometry solution will drift as error accrues over time. This drift can make data association highly ambiguous, especially under nonlinear measurement-modalities, such as vision [6]. Because a single incorrect association can be catastrophic, such systems in practice must

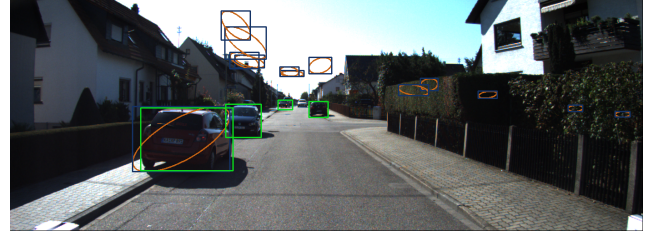


Fig. 1: SLAM algorithms are often presented with difficult data association tasks, especially after returning from long loops. Measurements generated by an object detector are shown as green bounding boxes, and the projections of estimated ellipsoidal landmarks are shown in orange, with predicted bounding boxes in blue. Rather than (potentially incorrectly) attempting to associate current measurements to existing landmarks, it is always safe to initialize a new landmark (e.g. car on left). Constellations of these duplicates can be merged together in a delayed fashion, providing a form of “lazy” data association.

use conservative recognition thresholds, choosing instead to initialize a new landmark whenever current measurements are poorly explained by the current set of landmark estimates. This conservatism is always “safe”, in the sense that the attribution of data to a new “duplicate” landmark will not cause estimator divergence, but it does come at the cost of increased model complexity and the loss of a valuable loop-closure constraint. Given that some level of front-end conservatism is unavoidable for robustness on real-world data, we propose a method for identifying and merging these duplicates in the SLAM back-end.

This paper presents a robust and efficient framework for efficient map-merging that is well-suited for semantic object-based SLAM. In order to achieve high precision in detecting duplicate landmarks, our approach identifies maximum-cardinality *constellations* of landmarks, sharing many similarities with joint gating techniques [7]–[9]. However, most existing methods assume constant-time access to the SLAM covariance matrix, which in recent nonlinear approaches to SLAM [10], [11] is not maintained explicitly and is expensive to recover.

To accomplish this, we define a geometric compatibility (GC) cost between two candidate constellations, capturing the most important correlations between landmarks but avoiding the reliance on global uncertainty information represented by the full covariance matrix. Furthermore, we introduce a conservative approximation of *local* landmark uncertainty that is incrementally-computable and facilitates efficient (linear-time) evaluation of the GC. Using GC-

Supported by the Defense Advanced Research Projects Agency (DARPA) as part of the Fast Lightweight Autonomy (FLA) program, HR0011-15-C-0110. Views expressed here are those of the authors, and do not reflect the official views or policies of the Dept. of Defense or the U.S. Government.

<sup>1</sup>K. Frey and J. How are with the Department of Aeronautics and Astronautics, MIT, Cambridge, MA 02139, USA. {kfrey, jhow}@mit.edu

<sup>2</sup>T. Steiner is a Senior Member of the Technical Staff at Draper, Cambridge, MA 02139, USA. tsteiner@draper.com

based search to eliminate the vast majority of hypotheses, we can restrict estimation of the JC (and requisite global uncertainty information) to only the most likely candidates (as an optional verification step).

As a secondary contribution, we propose a reformulation of the JC search as a set inclusion problem over a correspondence graph. In contrast to the traditional interpretation tree [7], [9], [12], we employ a more flexible binary search tree (BST) that is naturally constrained to cliques on the correspondence graph (which encode satisfaction of unary and binary constraints). This facilitates a stronger bounding for the branch-and-bound maximization, resulting in a significantly-accelerated optimization.

Our GC metric is verified on synthetic data with varying levels of noise, demonstrating desirable statistical properties in spite of nonlinear observations and significant estimate drift. Our approach achieves comparable performance to JC search (which requires global covariance information) at a fraction of the computational cost.

## II. RELATED WORK

The map merging and data association in this paper touches many subfields of SLAM (and robotics in general). As a solution to the loop-closure and relocalization problem, it provides an alternative to direct image-based localization methods, such as [13]–[15]. However, such systems require a rich database of “places” (i.e., keyframes) to localize against, and are sensitive to variations in viewpoint, lighting, or scene change. Recently, [16] achieved good robustness to extreme viewpoint and appearance variation by leveraging semantic information (similar in spirit to our approach) via per-pixel semantic classification.

Data association in SLAM is a well-explored problem, with solutions in applications ranging from acoustic sensing [17], [18], monocular vision [19], and LIDAR [20]. Traditionally, the search for jointly-compatible hypotheses has been approached as a search over the interpretation tree [12], [21] or a max-clique problem over a correspondence graph [22], [23]. The probabilistic Joint Compatibility (JC) metric introduced by [7] has been widely used as the de facto standard in filtering approaches (in which the full covariance matrix is readily available). A number of hybrid approaches [8] have been proposed, generally leveraging the correspondence graph to generate pairwise-compatible [12], [24] hypotheses and using JC to verify them [25]. In the case of feature cloud matching, which applies directly to sensor modalities such as LIDAR, [9], [26] leverage the specific independence structure to provide linear and constant-time incrementalized evaluations of JC. In contrast to the traditional gating formulation, the equivalent posterior form of the test is exploited in [27] that conveniently decouples different components of the error, allowing sequential computation.

Various other (non-gating) approaches to joint data association exist, including RANSAC [28], voting schemes [29], and explicit maximum-likelihood search [30]. Alongside these, several robust back-end approaches exist to identify and disable outlier measurements, based on residual gating

[31], linear programming [32], Expectation-Maximization [33], and explicit integer optimization [34]. Our approach is complementary to these, in that it adds a “second-stage” data-association, generating high-confidence candidate loop-closures which can be further verified by a robust back-end.

More along the lines of this work, [35] merge “duplicate” landmarks based on a semantic-aware clustering algorithm. However, the clustering formulation is somewhat limited, as data associations (i.e. merge decisions) are made individually rather than jointly.

## III. PRELIMINARIES

The goal of SLAM is to estimate a set of poses  $\mathcal{T} = \{T_i\}$  and landmarks  $\mathcal{L} = \{L_j\}$  given a set of noisy sensor measurements  $\{z_k\}$ . Each pose is a rigid-body transform  $T_i = (\mathbf{R}_i, \mathbf{t}_i)$ , with  $\mathbf{R}_i \in \text{SO}(3)$  and  $\mathbf{t}_i \in \mathbb{R}^3$  the rotation and translation components, respectively. Notationally, we will use leading superscripts when referring with respect to a specific coordinate system, e.g.  ${}^wT_i$  refers to the pose with respect to the global frame  $w$ . Because we are interested particularly in the *semantic* variant of SLAM, each landmark  $L_j$  may represent not just a point position  $\mathbf{p}_j \in \mathbb{R}^3$  in space but also a class label  $c_j$  and appearance descriptor  $\Omega_j$ . In the authors’ experience, state-of-the-art object detectors [1], [2] achieve good classification accuracy, and to simplify the discussion in this paper we assume class labels  $\{c_j\}$  are known accurately and can be leveraged as a hard constraint when identifying duplicate landmarks (see Sec. III-C).

In particular, we focus on “relative” SLAM problems, in which globally-referenced measurements (i.e. from GPS) are unavailable, and the estimation problem is only defined up to an arbitrary navigation frame  $w$ . Specifically, this means that observation factors  $\phi_{\text{obs}}$  and odometry factors  $\phi_{\text{odom}}$  are functions that can be expressed

$$\begin{aligned}\phi_{\text{obs}}({}^wT_i, {}^w\mathbf{p}_j) &= f({}^w\mathbf{R}_i^T({}^w\mathbf{p}_j - {}^w\mathbf{t}_i)) \\ \phi_{\text{odom}}({}^wT_i, {}^wT_{i+1}) &= g({}^w\mathbf{R}_i^T({}^w\mathbf{t}_{i+1} - {}^w\mathbf{t}_i), {}^w\mathbf{R}_i^T {}^w\mathbf{R}_{i+1})\end{aligned}\quad (1)$$

The resultant probabilistic estimation problem can be visualized as a factor graph [10] with variables nodes representing the quantities to be estimated ( $\mathcal{T}, \mathcal{L}$ ), and factor nodes representing the measurement models and sensor data relating them.

SLAM problems can often be quite large, involving hundreds of poses and thousands of landmarks. Modern smoothing approaches [10], [11] capitalize on the natural sparsity of such systems to perform efficient non-linear inference. This is accomplished in part by avoiding computation of the fully dense covariance matrix  ${}^w\Sigma$ , a contrast to traditional filtering approaches.  ${}^w\Sigma$  can be recovered at any time, but requires a large (and computationally expensive)  $n \times n$  matrix inversion (where  $n$  is scalar dimension of the system).

While recovering the full  ${}^w\Sigma$  is computationally intractable for many SLAM applications, the minimal “query” required to evaluate a constellation match hypothesis involves only a limited sub-block. Computing these marginal sub-blocks in general requires a partial Gaussian elimination over the factor graph, followed by a relatively small matrix

inversion, and efficient algorithms have been proposed [36]. If the number and size of such queries is limited to only the most promising hypotheses, this computation can be affordable in practice.

#### A. Delayed Data Association as Map-Merging

Data association can be viewed as the problem of finding the optimal mapping between a set of  $m$  measurements and  $n$  estimated landmarks, where it is possible for some measurements to be spurious, or to arise from previously-unseen “new” landmarks. In our context of map-merging, we attempt to find correspondences from *landmarks* to landmarks.

Notationally, we consider *candidate matches* to be ordered pairs  $s = (a, b)$  where  $a \neq b \in \{1, 2, \dots, n\}$  are the indices of two estimated landmarks  $L_a, L_b \in \mathcal{L}$ . The ordering of these indices is significant, because it implies that a candidate *merge set*  $\mathcal{C} = \{s_k\}$  is composed of two well-defined “constellations”  ${}^A\mathcal{C} = \{a : (a, b) \in \mathcal{C}\}$  and  ${}^B\mathcal{C} = \{b : (a, b) \in \mathcal{C}\}$ . For convenience, we will also assume that merge sets are ordered. Thus we can index into  ${}^A\mathcal{C} = (a_1, a_2, \dots, a_m)$  and  ${}^B\mathcal{C} = (b_1, b_2, \dots, b_m)$ .

#### B. Joint Compatibility

First we introduce a landmark-to-landmark variant of the joint compatibility (JC) criteria introduced by [7]. Replacing the “observation model” in feature-to-landmark association, define the matching residual between any two landmarks  $(L_a, L_b)$

$$\mathbf{r}(s) \triangleq \mathbf{p}_a - \mathbf{p}_b \quad (2)$$

Under the JC framework, this residual is evaluated in the *global* frame, and thus statistically depends on the global-frame covariance  ${}^w\Sigma$ . For a given hypothesis  $\mathcal{C} = ({}^A\mathcal{C}, {}^B\mathcal{C})$  of cardinality  $m$ , define a stacked residual  ${}^w\mathbf{r}^T = [{}^w\mathbf{r}(s_1)^T, {}^w\mathbf{r}(s_2)^T, \dots, {}^w\mathbf{r}(s_m)^T]$  and corresponding covariance

$${}^w\Sigma_{\mathbf{r}} = {}^w\Sigma_{{}^A\mathcal{C}} + {}^w\Sigma_{{}^B\mathcal{C}} - {}^w\Sigma_{{}^A\mathcal{C}, {}^B\mathcal{C}} - {}^w\Sigma_{{}^B\mathcal{C}, {}^A\mathcal{C}} \quad (3)$$

where  ${}^w\Sigma_{{}^A\mathcal{C}}$ ,  ${}^w\Sigma_{{}^B\mathcal{C}}$ , and  ${}^w\Sigma_{{}^A\mathcal{C}, {}^B\mathcal{C}}$  are the corresponding  $3m \times 3m$  sub-blocks of  ${}^w\Sigma$ .

The JC criterion can be expressed

$${}^w\mathbf{r}^T {}^w\Sigma_{\mathbf{r}}^{-1} {}^w\mathbf{r} < d_{\chi^2, 3m} \quad (4)$$

It should be noted that using JC as a search criteria over the combinatoric set of joint hypotheses over all  $\mathcal{O}(n^2)$  candidate matches requires computing (3) and therefore the full  ${}^w\Sigma$ .  ${}^w\Sigma$  can of course be pre-computed before starting the search, but this in general must be repeated at each time step (as loop-closure or linearization point updates can affect covariances globally). Furthermore, the presence of nonlinearities in the SLAM system can result in overconfidence in the linearized uncertainty estimate. For these reasons, we replace the JC criterion with a geometric compatibility (GC) condition that avoids this dependence on global covariance information, facilitating computationally-lightweight and accurate gating.

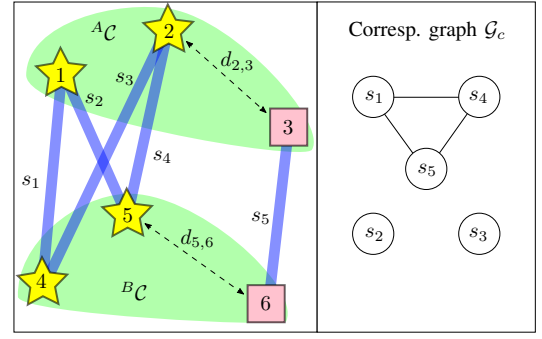


Fig. 2: Drift in translation and rotation manifests as “duplicate” landmark constellations  ${}^A\mathcal{C}$  and  ${}^B\mathcal{C}$  (left). Class labels are drawn as either a star or square, and unary-compatible matches  $\{s_i\}$  are shown in blue. A necessary condition for joint compatibility is preservation of distances, e.g. that  $d_{2,3} \approx d_{5,6}$ . This pairwise “rigidity” criterion can be formulated as a binary constraint between pairs  $(s_i, s_j)$  as in [12]. Satisfaction of all binary constraints between candidates  $(s_i, s_j)$  induces an edge in  $\mathcal{G}_c$  (right), and pairwise-compatible hypotheses form cliques.

#### C. Pairwise Compatibility over a Correspondence Graph

Fundamental to our approach is a tight integration between explicit tree search and clique-based compatibility. Given a set of unary and binary constraints on candidate matches and pairs of matches, respectively, a correspondence graph [22], [24] can be defined. Candidates  $s_i$  satisfying unary constraints are represented as nodes, with edges connecting pairs  $(s_i, s_j)$  that satisfy binary constraints. Thus, cliques on this graph represent *pairwise*-compatible sets of candidates (a weaker condition than joint compatibility). The unary and binary constraints in question are generic, and can represent geometric constraints [12], [24], locality [8], appearance and class similarity, or other expert knowledge.

If it is assumed that these constraints are in fact *sufficient* for joint compatibility, the search problem can be formulated as a max-clique problem [22]–[24]. While this has the benefit of taking advantage of off-the-shelf graph-theoretic algorithms, the sufficiency assumption can be limiting. Nevertheless, this correspondence graph can be much more efficient than naive JC tree search, as the binary constraints effectively eliminate entire branches of the tree, and are evaluated only once for each pair  $(s_i, s_j)$ . As explained in Section V, the correspondence graph provides a hitherto unexploited tight upper bound for max-cardinality maximization that can significantly accelerate the search.

A number of unary and binary constraints may apply to the scenario of semantic map-merging. The minimal set of constraints assumed in this paper is given below.

**Unary:**  $s = (a, b)$

(U<sub>1</sub>) Disjoint:  $a$  not equal to  $b$

(U<sub>2</sub>) Class label match:  $c_a = c_b$

(U<sub>3</sub>) Min separation (Sec. IV-B)

**Binary:**  $s_1 = (a_1, b_1)$ ,  $s_2 = (a_2, b_2)$

(B<sub>1</sub>) Disjoint:  $a_1, b_1, a_2$  and  $b_2$  are distinct indices

(B<sub>2</sub>) Locality:  $(a_1, a_2)$  “close”,  $(b_1, b_2)$  “close” (Sec. IV-B)

#### D. Max-Cardinality Search

In practice, we wish to find the largest set of correspondences that satisfies a suite of compatibility conditions. As an optimization, this can be formulated as a max-cardinality search over cliques  $\mathcal{C}(\mathcal{G}_c)$  in our correspondence graph  $\mathcal{G}_c$ .

$$\max_{\mathcal{C} \in \mathcal{C}(\mathcal{G}_c)} |\mathcal{C}| \quad (5)$$

$$\text{subject to: } \text{Compatible}(\mathcal{C}) \quad (6)$$

Without the compatibility constraint (6), this reduces to a max-clique search, as in [22], but with it an explicit tree search is required.

The combinatoric search over all possible assignments  $\mathcal{C}$  has traditionally been visualized as an interpretation tree [7], [12]. The interpretation tree has a branching factor of  $n + 1$  (the number of matchable landmarks plus a “null” match), and a depth of  $m$  (the number of “measurements”). The path from the root to any leaf node describes a potential joint assignment  $\mathcal{C}$ , and the structure of the tree compactly imposes the constraint that a single measurement cannot match to more than one landmark. At each step of the search, a partial hypothesis is evaluated, and if the constraint (6) is not satisfied, the algorithm discontinues exploration of the corresponding branch. Thus, a strong compatibility metric (i.e. JC) is vital to efficiently prune the search.

Though standard, the interpretation tree framework (and associated algorithms) has a significant weakness: it cannot efficiently represent the requirement that  $\mathcal{C} \in \mathcal{C}(\mathcal{G}_c)$ . Though unary and binary constraints can of course be checked at each step of the tree search [8], this is inherently inefficient because the same candidate matches  $s_i, s_j$  will be tested multiple times. This fact has historically made max-clique and tree-search approaches largely disparate, with “hybrid” algorithms essentially switching from clique-based hypothesis generation to tree-based verification [8]. Additionally, efficient branch-and-bound maximization requires the availability of tight upper bounds. However, the main feasibility criteria  $\mathcal{C} \in \mathcal{C}(\mathcal{G}_c)$  cannot be directly “read” from the interpretation tree, and thus much weaker bounds based solely on tree depth are used in practice [19, Alg. 2]. In Section V, we reformulate the interpretation tree as a set inclusion problem over a binary-search-tree (BST), a more flexible framework allowing tighter (and therefore more efficient) branch-and-bound search.

#### IV. GEOMETRIC COMPATIBILITY AND LOCAL UNCERTAINTY

Assume we are given two constellations  ${}^A\mathcal{C}$  and  ${}^B\mathcal{C}$  of cardinality  $m > 1$  which are well-separated in the graph (i.e. that estimate correlations are small between them, as might be the case when a robot returns from a long loop). If the two constellations can be considered “locally rigid” (a concept which will be made more precise in the following section), the JC error can be thought of as simultaneously capturing geometric error (how well constellations “match” under optimal alignment) and drift error (distance in translation and rotation). Importantly, the geometric error is a local

property (involving only the local subgraphs of  ${}^A\mathcal{C}$  and  ${}^B\mathcal{C}$  respectively) while the drift error is a global property of the posterior distribution. To exploit this fact, we introduce a convenient approximation of local uncertainty that decouples the geometric error from the rest of the graph. This decoupling allows efficient search for maximum-cardinality joint hypotheses  $\mathcal{C}$  satisfying this GC criterion, which can then be verified globally via a JC test as an optional post-step.

##### A. Geometric Compatibility (GC)

In order to decouple the joint compatibility between constellations  ${}^A\mathcal{C} = \{a_1, a_2, \dots, a_m\}$  and  ${}^B\mathcal{C} = \{b_1, b_2, \dots, b_m\}$ , we consider the geometric fit given the *optimal* rigid-body alignment. In the general case, we then have two corresponding sets of landmark estimates,  $\{{}^A\mathbf{p}_{a_i}\}$  and  $\{{}^B\mathbf{p}_{b_i}\}$ , in distinct coordinate frames  $A$  and  $B$ . Following (2), the residual in frame  $A$  (given some alignment  ${}^AT_B$ ) is

$${}^A\mathbf{r}_i = {}^A\mathbf{p}_{a_i} - {}^AT_B \circ {}^B\mathbf{p}_{b_i} \quad (7)$$

and the geometric compatibility GC is defined

$$d_{GC} \triangleq \min_{{}^AT_B \in \text{SE}(3)} \mathbf{r}^T ({}^A\mathbf{\Sigma} + [{}^A\mathbf{R}_B]^B \mathbf{\Sigma} [{}^A\mathbf{R}_B]^T)^{-1} \mathbf{r} \quad (8)$$

where  $\mathbf{r}^T = [\mathbf{r}_1^T, \mathbf{r}_2^T, \dots, \mathbf{r}_m^T]$  is the stacked residual vector (the frame superscripts are dropped for clarity),  ${}^A\mathbf{\Sigma}$  and  ${}^B\mathbf{\Sigma}$  are the  $3m \times 3m$  covariance matrices in frames  $A$  and  $B$ , respectively, and  $[\mathbf{R}] \triangleq \text{BlockDiag}(\mathbf{R}, \mathbf{R}, \dots, \mathbf{R}) \in \text{SO}(3m)$  is a  $3m \times 3m$  block diagonal matrix. The resulting gating test  $d_{GC} < d_{\chi^2, 3m-6}$  is a *posterior* compatibility test [27] over  $3m-6$  degrees of freedom with the point estimates corresponding to  ${}^A\mathcal{C}$  and  ${}^B\mathcal{C}$  treated as independent sets of “measurements.” This independence can be safely assumed if  ${}^A\mathcal{C}$  and  ${}^B\mathcal{C}$  are sufficiently separated in the graph, which can be enforced via  $(U_3)$ .

Note that evaluating GC involves a nonlinear optimization over  $\text{SE}(3)$ . We achieve an efficient approximation in a two-step approach. First, we approximate (8) with an orthogonal Procrustes optimization [37]

$$\bar{T} = \underset{T \in \text{SE}(3)}{\text{argmin}} \mathbf{r}^T \mathbf{r} = \underset{T \in \text{SE}(3)}{\text{argmin}} \sum_{j=1}^m \|{}^A\mathbf{p}_{a_j} - T \circ {}^B\mathbf{p}_{b_j}\|_2^2. \quad (9)$$

whose solution  $\bar{T}$  can be computed in closed-form. Then, we refine this estimate on the full objective function (8) with a tangent-space linearization  $\mathbf{r}_i = {}^A\mathbf{r}_{a_i} - (\bar{T} \circ \text{Exp}(\delta)) \circ {}^B\mathbf{p}_{b_i}$  where  $\delta \in \mathbb{R}^6$ . By “locking” the covariance terms, this produces a small linear-least-squares problem over  $\delta$  that can be solved efficiently.

In contrast to the IPJC algorithm [9], our approach operates without a prior on  ${}^AT_B$ , which would require access to  ${}^w\mathbf{\Sigma}$ . In practice, we find that our two-step optimization provides a suitable approximation of (8) in constant-time.

##### B. Approximating Local Uncertainty

In order to compute (8), we need estimates of  ${}^A\mathbf{\Sigma}$  and  ${}^B\mathbf{\Sigma}$  in some to-be-determined frames  $A$  and  $B$ . To simplify the following discussion, define  $\mathcal{T}_j \subset \mathcal{T}$  to be the set of poses from which landmark  $L_j$  is observed. Furthermore,



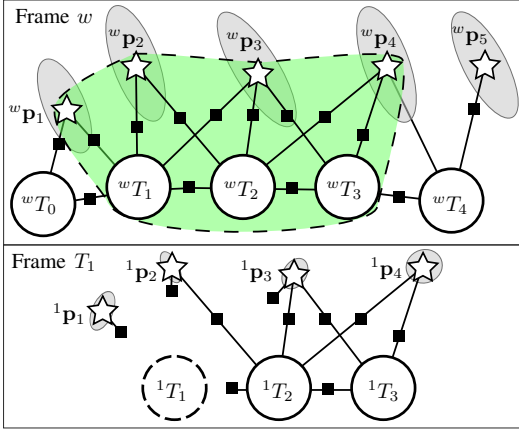


Fig. 3: [top] Example SLAM graph expressed in the world frame. The green shaded region indicates the local subgraph extracted to approximate local uncertainty over  $\mathbf{p}_3$ . [bottom] This subgraph expressed in local frame of  $T_1$ . Note that marginal covariances (shaded ellipses) over the landmarks in this frame are generally smaller than in the world frame and show less correlation.

assume that  $\mathcal{T}_{AC} \triangleq \bigcap_{i=1}^m \mathcal{T}_{a_i} \neq \emptyset$  and  $\mathcal{T}_{BC} \triangleq \bigcap_{i=1}^m \mathcal{T}_{b_i} \neq \emptyset$  – that is that there exists at least one common pose adjacent to all of  ${}^A\mathcal{C}$  and another (distinct) pose adjacent to all of  ${}^B\mathcal{C}$ . In the case that  $\mathcal{T}_m$  are *intervals*, this can be enforced efficiently during search with a pairwise locality constraint ( $B_2$ ). Nevertheless, the approach outlined here can be straightforwardly extended to more general scenarios.

One approach could be to select a  $T_A \in \mathcal{T}_{AC}$  and  $T_B \in \mathcal{T}_{BC}$  and compute  ${}^A\Sigma$  and  ${}^B\Sigma$  “on-the-fly” during search. However, because each landmark is eventually considered as part of numerous constellations, this can lead to significant redundant computation. Instead, we seek to pre-compute a set of *independent* marginals for each  $\mathbf{p}_j$ , one for each local frame represented in  $\mathcal{T}_j$ .

Our process is illustrated in Fig. 3 for a given  $L_j$ . We first extract the local subgraph containing  $\mathbf{p}_j$ ,  $\mathcal{T}_j$ , and all landmarks adjacent to  $\mathcal{T}_j$ . For each  $T_i \in \mathcal{T}_j$  we compute the  $3 \times 3$  marginal  ${}^i\Sigma_{jj}$  over  ${}^i\mathbf{p}_j$  in the corresponding local frame. This computation involves only the local subgraph, and assuming a constant max cardinality  $|\mathcal{T}_j| \leq N$ , this can be accomplished in constant time for each  $L_j$ . Furthermore, because it involves only local information (and is independent of global linearization point), it can be performed incrementally (as only recent landmarks will need to be updated). Finally, because some information is ignored (specifically, observations of other local landmarks), this estimate is conservative.

It should be noted that for nearby landmarks, the resulting estimates *are* correlated, although in practice we’ve found these correlations to be small and the independence approximation to be sufficient in light of the computational advantages. In some scenarios (specifically pairwise comparisons), the effect of these correlations can be explicitly bounded, although for brevity further exploration is omitted here.

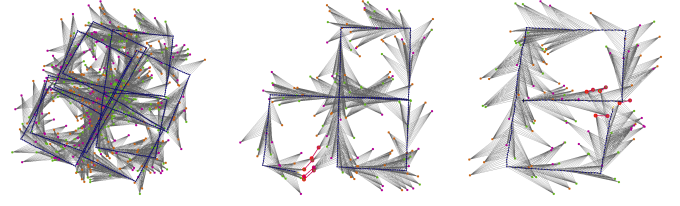


Fig. 4: Constellation merging in simulation. The robot ground-truth trajectory happens to be planar, but the state space is fully 3-dimensional. [left] Without constellation merging, localization error increases over time, and the final estimate shows significant error as well as many duplicate landmarks. [center, right] Correct constellation matches are detected despite significant drift in rotation and translation.

### C. Linear-complexity GC

As stated before, we cache a set of marginals  $\{{}^i T_{jj}\}$  for each landmark  $\mathbf{p}_j$ . During evaluation of (8), the sets of common frames  $\mathcal{T}_{AC}$  and  $\mathcal{T}_{BC}$  are assured to be non-empty, and thus  $T_A \in \mathcal{T}_{AC}$  and  $T_B \in \mathcal{T}_{BC}$  can be selected via min-determinant or min-trace criteria.

By assuming independence between each  $\mathbf{p}_j$ ,  ${}^A\Sigma$  and  ${}^B\Sigma$  become block-diagonal, and (8) simplifies to a linear sum-of-squares

$$d_{GC} = \sum_{j=1}^m \mathbf{r}_j^T ({}^A\Sigma_{jj} + {}^A\mathbf{R}_B {}^B\Sigma_{jj} {}^A\mathbf{R}_B^T)^{-1} \mathbf{r}_j \quad (10)$$

Thus, the independence assumption allows linear-time evaluation of GC, compared to the generally quadratic evaluation of (8).

### V. RE-INTERPRETING THE INTERPRETATION TREE

Rather than defining hypotheses over an interpretation tree [12], we explicitly approach the maximization (5) as a set inclusion problem over  $\mathcal{C}(\mathcal{G}_c) \subset 2^{m \times n}$ . This is represented as a binary search tree (BST), where each level of the BST corresponds to the inclusion or disclusion of a unary-feasible candidate  $s_i \in \mathcal{V}(\mathcal{G}_c)$ . Every node in the tree (say at depth  $d$ ) corresponds to a partial hypothesis  $\mathcal{C}_d$  in which only  $d \leq |\mathcal{V}(\mathcal{G}_c)|$  candidates have been considered. Critically, inclusion of the next vertex  $s_i$  is only considered if  $s_i$  is adjacent to every previously included node  $s_j$  in  $\mathcal{C}_d$ . This enforces that  $\mathcal{C}_d \in \mathcal{C}(\mathcal{G}_c)$  always, and can be implemented efficiently by maintaining a list of remaining unconsidered, but jointly-adjacent, nodes  $S(\mathcal{C}_d) \subset \mathcal{V}(\mathcal{G}_c)$  for each partial hypothesis  $\mathcal{C}_d$ . Given  $\mathcal{C}_d$  and  $S(\mathcal{C}_d)$ , a tight upper bound is

$$\forall \mathcal{C} \in \mathcal{C}(\mathcal{G}_c) : \mathcal{C} \supset \mathcal{C}_d \text{ we have } |\mathcal{C}| \leq \text{UpperBound}(\mathcal{C}_d) = |\mathcal{C}_d| + |S(\mathcal{C}_d)|. \quad (11)$$

This BST approach is clearly more general than the interpretation tree, as it does *not* implicitly enforce the constraint that each measurement corresponds to at most one landmark. However, this can be easily re-imposed via the binary disjointness constraint ( $B_1$ ). When  $\mathcal{G}_c$  reflects *only* this disjointness constraint, the BST and interpretation tree approaches are identical. However, in the presence of other constraints, the BST is superior in that it only tests

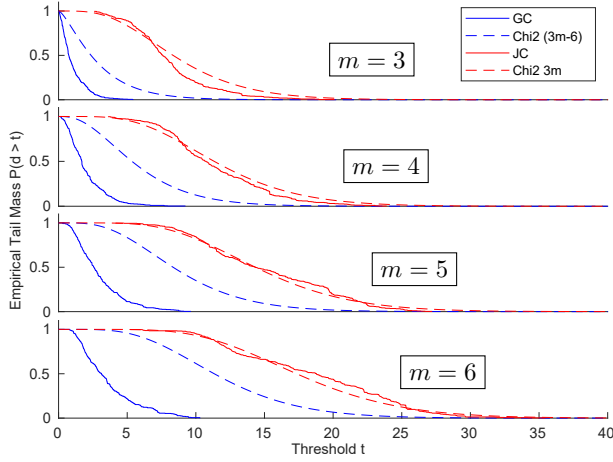


Fig. 5: Empirical compatibility histograms (solid) vs. corresponding  $\chi^2$  tail mass (dashed) for true constellation matches of varying cardinalities. For the most part the tail mass of the JC metric follows the prediction, but shows some discrepancy due to global-frame nonlinearity. As expected, our conservative GC estimates decay faster than the corresponding  $\chi^2$  tail, ensuring that we do not “miss” good matches.

each candidate  $s$  and pair of candidates  $(s_i, s_j)$  *once* (while building  $\mathcal{G}_c$ ), and can leverage these constraints to provide a tighter upper bound via (11). In doing so, it unifies clique- and tree-based search schemes in a straightforward, easily-implemented way.

## VI. EXPERIMENTAL RESULTS

The proposed statistics and methods were validated in a simulated nonlinear visual-SLAM setting implemented with GTSAM [38]. As the robot moves along a loopy trajectory shown in Fig. 4, it receives noisy odometry and makes observations of randomly-distributed landmarks via a simulated single-camera sensor, with limited range and field-of-view. To simulate a “short-term” data association solution (e.g. frame-to-frame tracking), landmark associations are “lost” once the feature leaves the camera field of view, and further observations are assigned to a new, duplicate landmark. Thus, if no map-merging is performed, there is no global loop-closure, and localization uncertainty grows with time.

To emulate semantic SLAM, landmarks are randomly assigned one of three classes (indicated by color), and it is assumed that class label is accurately observed (reasonable given the performance of state-of-the-art detectors [1]). This semantic information is used as a unary constraint ( $U_2$ ) to help sparsify the correspondence graph.

Fig. 5 demonstrates the statistical consistency of our GC metric on randomly-sampled *ground-truth* constellation matches in simulation. Because of lossy linearizing approximations, the JC metric does not perfectly follow a  $\chi^2$  distribution, whereas the GC scores are (correctly) conservative. This conservatism arises from the practical need to estimate landmark uncertainties using only a limited subset of the available data (see Sec. IV-B). Here we use a minimal subset, although larger subsets could be chosen (at the cost

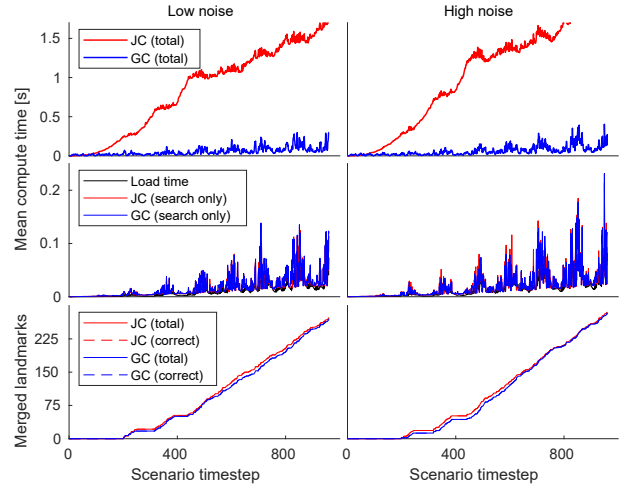


Fig. 6: Averaged simulation results over low and high noise conditions, comparing joint-compatibility to our method. Both methods identify comparable numbers of merges (bottom), and achieve over 99% accuracy in all tests. The  ${}^w\Sigma$  pre-computation required by JC dominates total computation times (top), while the time spent in actual tree search (middle) is similar for both methods. Note the difference in axis scales. The time spent updating the correspondence graph  $\mathcal{G}_c$  (common to both methods) is shown in black.

of more computation). Fig. 6 shows a comparison in both detection performance and computation time between the proposed GC metric and baseline JC [7]. As can be seen, a similar number of matches are found using the GC, but without the expensive step of computing  ${}^w\Sigma$ . Thus, our GC method performs as well as a full JC-based search but at a fraction of the computational cost.

## VII. CONCLUSIONS

Measurements are most informative when the estimate has drifted, but that is when they are simultaneously the most ambiguous. Given the catastrophic risks of incorrect associations, it is always safer to ascribe ambiguous measurements to a new landmark than to an existing one. With this principle in mind, this paper introduces an efficient method of “delayed” data association via landmark constellation merging. While most relevant methods assume access to the full covariance matrix and/or fully uncorrelated measurements (i.e. in feature cloud matching), our method leverages local and incrementally-computable information to identify good candidates over the full graph. If needed, the sparse set of resulting matches can then be verified via standard covariance-based methods at a computational cost that is tenable in practice. We believe that our GC-based approach provides a robust, secondary level of loop-closure detection *in the back-end* that facilitates the re-capture of “missed” loop closures, reducing the burden on front-end data association.

## ACKNOWLEDGMENT

Thanks to Dr. Kasra Khosoussi for the many productive conversations and input over the course of this work.

## REFERENCES

- [1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comp. Vision Pat. Rec. (CVPR)*, 2016, pp. 779–788.
- [2] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. European Conf. on Comp. Vision (ECCV)*. Springer, 2016, pp. 21–37.
- [3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Conf. on Neural Inf. Proc. Systems (NeurIPS)*, 2015, pp. 91–99.
- [4] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," vol. 32, no. 6, pp. 1309–1332, 2016.
- [5] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," *Proc. European Conf. on Comp. Vision (ECCV)*, pp. 404–417, 2006.
- [6] L. J. Nicholson, M. J. Milford, and N. Sunderhauf, "QuadricSLAM: Dual quadrics from object detections as landmarks in object-oriented SLAM," *IEEE Robotics and Automation Letters*, 2018.
- [7] J. Neira and J. D. Tardós, "Data association in stochastic mapping using the joint compatibility test," *IEEE Trans. Robot. Automat.*, vol. 17, no. 6, pp. 890–897, 2001.
- [8] J. Neira, J. D. Tardós, and J. A. Castellanos, "Linear time vehicle relocation in SLAM," in *Proc. IEEE Conf. Robot. Autom. (ICRA)*. Citeseer, 2003, pp. 427–433.
- [9] Y. Li and E. B. Olson, "IPJC: The incremental posterior joint compatibility test for fast feature cloud matching," in *Proc. IEEE Conf. Int. Rob. Sys. (IROS)*. IEEE, 2012, pp. 3467–3474.
- [10] F. Dellaert and M. Kaess, "Square root SAM: Simultaneous localization and mapping via square root information smoothing," *Int. J. of Robotics Research*, vol. 25, no. 12, pp. 1181–1203, 2006.
- [11] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: incremental smoothing and mapping using the bayes tree," *Int. J. of Robotics Research*, 2011.
- [12] W. E. L. Grimson, D. P. Huttenlocher, et al., *Object recognition by computer: the role of geometric constraints*. MIT Press, 1990.
- [13] P. Newman and K. Ho, "SLAM-loop closing with visually salient features," in *Proc. IEEE Conf. Robot. Autom. (ICRA)*. IEEE, 2005, pp. 635–642.
- [14] R. Paul and P. Newman, "FAB-MAP 3D: Topological mapping with spatial and visual appearance," in *Proc. IEEE Conf. Robot. Autom. (ICRA)*. IEEE, 2010, pp. 2649–2656.
- [15] B. Williams, G. Klein, and I. Reid, "Automatic relocation and loop closing for real-time monocular SLAM," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 33, no. 9, pp. 1699–1712, 2011.
- [16] S. Garg, N. Sunderhauf, and M. Milford, "LoST? appearance-invariant place recognition for opposite viewpoints using visual semantics," in *Robotics Science and Systems*, vol. 14, 2018.
- [17] T. A. Huang and M. Kaess, "Towards acoustic structure from motion for imaging sonar," in *Proc. IEEE Conf. Int. Rob. Sys. (IROS)*. IEEE, 2015, pp. 758–765.
- [18] M. F. Fallon, J. Folkesson, H. McClelland, and J. J. Leonard, "Relocating underwater features autonomously using sonar-based SLAM," *IEEE Journal of Oceanic Engineering*, vol. 38, no. 3, pp. 500–513, 2013.
- [19] L. A. Clemente, A. J. Davison, I. D. Reid, J. Neira, and J. D. Tardós, "Mapping large loops with a single hand-held camera," in *Robotics Science and Systems*, vol. 2, no. 2, 2007.
- [20] J. E. Guivant and E. M. Nebot, "Optimization of the simultaneous localization and map-building algorithm for real-time implementation," *IEEE transactions on robotics and automation*, vol. 17, no. 3, pp. 242–257, 2001.
- [21] D. Hähnel, S. Thrun, B. Wegbreit, and W. Burgard, "Towards lazy data association in SLAM," in *Robotics Research. The Eleventh International Symposium*. Springer, 2005, pp. 421–431.
- [22] T. Bailey, E. M. Nebot, J. Rosenblatt, and H. F. Durrant-Whyte, "Data association for mobile robot navigation: A graph theoretic approach," in *Proc. IEEE Conf. Robot. Autom. (ICRA)*, vol. 3. IEEE; 1999, 2000, pp. 2512–2517.
- [23] P. San Segundo and D. Rodriguez-Losada, "Robust global feature based data association with a sparse bit optimized maximum clique algorithm," *IEEE Transactions on Robotics*, vol. 29, no. 5, pp. 1332–1339, 2013.
- [24] J. H. Lim, J. J. Leonard, and S. K. Kang, "Mobile robot relocation using echolocation constraints," in *Proc. IEEE Conf. Int. Rob. Sys. (IROS)*, vol. 1. IEEE, 1999, pp. 154–159.
- [25] L. M. Paz, J. D. Tardós, and J. Neira, "Divide and conquer: EKF SLAM in  $O(n)$ ," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1107–1120, 2008.
- [26] X. Shen, E. Frazzoli, D. Rus, and M. H. Ang, "Fast joint compatibility branch and bound for feature cloud matching," in *Proc. IEEE Conf. Int. Rob. Sys. (IROS)*. IEEE, 2016, pp. 1757–1764.
- [27] Y. Li, S. Li, Q. Song, H. Liu, and M. Q.-H. Meng, "Fast and robust data association using posterior based approximate joint compatibility test," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 331–339, 2014.
- [28] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [29] L. M. Paz, P. Piniés, J. Neira, and J. D. Tardós, "Global localization in SLAM in bilinear time," in *Proc. IEEE Conf. Int. Rob. Sys. (IROS)*. IEEE, 2005, pp. 2820–2826.
- [30] N. Atanasov, M. Zhu, K. Daniilidis, and G. J. Pappas, "Semantic localization via the matrix permanent," in *Robotics Science and Systems*, vol. 2, 2014.
- [31] M. C. Graham, J. P. How, and D. E. Gustafson, "Robust incremental SLAM with consistency-checking," in *Proc. IEEE Conf. Int. Rob. Sys. (IROS)*. IEEE, 2015, pp. 117–124.
- [32] L. Carlone, A. Censi, and F. Dellaert, "Selecting good measurements via 1 relaxation: A convex approach for robust estimation over graphs," in *Proc. IEEE Conf. Int. Rob. Sys. (IROS)*. IEEE, 2014, pp. 2667–2674.
- [33] S. L. Bowman, N. Atanasov, K. Daniilidis, and G. J. Pappas, "Probabilistic data association for semantic SLAM," in *Proc. IEEE Conf. Robot. Autom. (ICRA)*. IEEE, 2017, pp. 1722–1729.
- [34] N. Sünderhauf and P. Protzel, "Switchable constraints for robust pose graph SLAM," in *Proc. IEEE Conf. Int. Rob. Sys. (IROS)*. IEEE, 2012, pp. 1879–1884.
- [35] B. Mu, S.-Y. Liu, L. Paull, J. Leonard, and J. P. How, "SLAM with objects using a nonparametric pose graph," in *Proc. IEEE Conf. Int. Rob. Sys. (IROS)*. IEEE, 2016, pp. 4602–4609.
- [36] M. Kaess and F. Dellaert, "Covariance recovery from a square root information matrix for data association," *Robotics and autonomous systems*, vol. 57, no. 12, pp. 1198–1210, 2009.
- [37] J. C. Gower, G. B. Dijkstra, et al., *Procrustes problems*. Oxford University Press on Demand, 2004, vol. 30.
- [38] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," Georgia Institute of Technology, Tech. Rep., 2012.