

# Regularized Graph Matching for Correspondence Identification under Uncertainty in Collaborative Perception

Peng Gao<sup>1</sup>, Rui Guo<sup>2</sup>, Hongsheng Lu<sup>2</sup>, and Hao Zhang<sup>1</sup>

<sup>1</sup>Colorado School of Mines, <sup>2</sup>Toyota Motor North America

gaopeng@mines.edu, rui.guo@toyota.com, hongsheng.lu@toyota.com, hzhang@mines.edu

**Abstract**—Correspondence identification is a critical capability for multi-robot collaborative perception, which allows a group of robots to consistently refer to the same objects in their own fields of view. Correspondence identification is a challenging problem, especially caused by non-covisible objects that cannot be observed by all robots and the uncertainty in robot perception, which have not been well studied yet in collaborative perception. In this work, we propose a principled approach of regularized graph matching that addresses perception uncertainties and non-covisible objects in a unified mathematical framework to perform correspondence identification in collaborative perception. Our method formulates correspondence identification as a graph matching problem in the regularized constrained optimization framework. We introduce a regularization term to explicitly address perception uncertainties by penalizing the object correspondences with a high uncertainty. We also design a second regularization term to explicitly address non-covisible objects by penalizing the correspondences built by the non-covisible objects. The formulated constrained optimization problem is difficult to solve, because it is not convex and it contains regularization terms. Thus, we develop a new sampling-based algorithm to solve our formulated regularized constrained optimization problem. We evaluate our approach in the scenarios of connected autonomous driving and multi-robot coordination in simulations and using real robots. Experimental results show that our method is able to address correspondence identification under uncertainty and non-covisibility, and it outperforms the previous techniques and achieves the state-of-the-art performance.

## I. INTRODUCTION

Multi-robot systems have been attracting a significant attention over the past decades due to their reliability, parallelism, and scalability to address large-scale problems [8, 15, 62]. As one of the essential abilities of multi-robot systems, collaborative perception enables shared awareness and understanding of the surrounding environment among the robots, which plays an important role in a variety of real-world applications, including robot-assisted search and rescue [36, 52, 53, 51], connected and autonomous driving [60], collaborative manufacture [17], and multi-robot localization and mapping [2, 48].

Correspondence identification is a core task in collaborative perception, with the goal of identifying the same objects (thus deciding the correspondences) observed by two robots in their own fields of view [21, 30, 39, 57, 22]. For example, as shown in Figure 1, if two connected vehicles want to share information of other vehicles and road conditions, they first need to identify street objects' correspondence in order to correctly refer to the same objects. In another example, if a

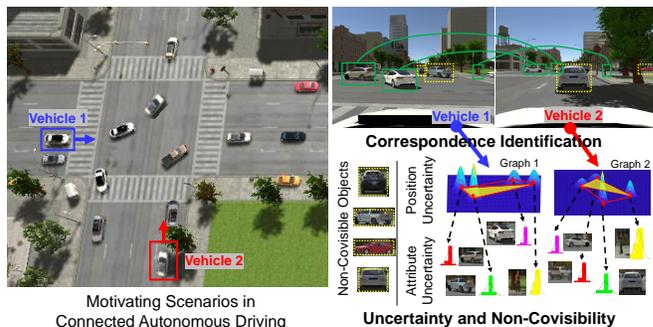


Fig. 1. A motivating example of correspondence identification with non-covisible objects under uncertainty for multi-robot collaborative perception. Before two connected autonomous vehicles share information of street objects, they must identify the correspondence of these objects, while addressing non-covisible objects and perception uncertainty, in order to correctly refer to the same objects from their own fields of view.

robot wants to acquire information of a target from another robot, correspondence must be identified to ensure that both robots refer to the same target. Due to the importance of correspondence identification, a variety of methods were recently implemented to address this problem. The first category of methods decide correspondence via calculating the appearance similarity of a pair of objects, e.g., based on visual features for appearance matching [13, 24]. The second category of methods employ synchronization algorithms to identify circle-consistent associations among objects [2, 20, 29, 58], which synchronize associations of the same objects observed from multiple views. The third category of methods are based on spatial and geometric information, e.g., using linear assignment [44], quadratic assignment [14, 40], and high-order graph matching [12, 18, 49].

Correspondence identification is a challenging task to solve in collaborative perception, because robots often observe a part of the environment from different views, and because multiple objects observed by these robots may look similar or identical. Specifically, although previous methods demonstrated encouraging results, two challenges in correspondence identification have not been well studied yet. This first challenge is resulted from *non-covisible objects*, which are those objects that cannot be observed by all robots, but are only observable by a subset of robots. Due to occlusion and robot's limited field of view,

non-covisible objects in collaborative perception are common. The non-covisible objects often greatly affect correspondence identification, since not all objects observed by multiple robots have a correspondence. The second challenge is resulted from *uncertainty* in perception. For example, attributes (e.g., visual or semantic features) extracted to describe objects can be noisy and ambiguous for correspondence identification. In addition, object positions (e.g., estimated by depth estimation algorithms [33]) used to compute spatial relationships of the objects often show a deviation from their real positions and are noisy.

In this paper, we introduce a novel regularized graph matching method to address the task of correspondence identification with non-covisible objects under uncertainty for collaborative perception. We use a graph representation to represent multiple objects observed by a robot. Each node denotes an object and is associated with an attribute vector, and the edges among the nodes are used to describe the spatial relationships among the objects. Then, we formulate correspondence identification as an optimization-based graph matching problem. In order to address non-covisible objects, we design a new regularization term over the number of identified correspondence to penalize the correspondence constructed by the non-covisible objects. In order to address uncertainty in the attributes and positions of the objects, we propose another regularization term to penalize the correspondence of objects with high attribute and position uncertainties and promote the correspondence of objects with low uncertainties.

The key contribution of this paper is the introduction of the first principled approach that addresses perception uncertainty and non-covisible objects in a unified mathematical framework to perform correspondence identification in collaborative perception. Specific novelties of the paper include:

- We introduce a novel regularized graph matching method for correspondence identification under perception uncertainty with non-covisible objects. Our approach integrates perception uncertainty into the graph representation, and uses two new regularization terms to reduce the influence resulted from uncertainty and non-covisibility.
- We implement an effective new optimization algorithm to solve the formulated constrained optimization problem that is challenging to solve as the problem is non-convex and contains two regularization terms.

As a practical contribution, we provide one of the first datasets in order to study the problem of correspondence identification with non-covisible objects in collaborative perception.

## II. RELATED WORK

### A. Correspondence Identification

Existing methods can be grouped into three categories based on appearance, synchronization, and spatial relationship.

*Appearance-based identification* identifies correspondences based on appearance similarities, which can be further divided into three subgroups using keypoints, visual appearances and semantic attributes, respectively. Keypoint-based methods are commonly used in matching adjacent frames in simultaneous

localization and mapping (SLAM) using key-points SIFT [19], ORB [45] and 3D keypoints [6]. Furthermore, to identify same individual objects with changing appearances, visual features [64] and attribute features [65] are used for re-identification.

*Synchronization-based identification* recognizes correspondences of objects from multiple views by satisfying the circle-consistent constraint [20]. Synchronization-based methods can be divided into three subgroups, based upon convex relaxation, spectral relaxation and graph clustering. The convex relaxation methods formulate multi-view correspondence identification as a semidefinite problem [7], which can be relaxed to be convex [29] and solved using a convex optimization solver [66]. The spectral relaxation methods also formulate it as a semidefinite problem and compute approximated solutions based upon top-rank eigen-vectors decomposed from the original formulation [43, 50]. The graph clustering methods formulate the multi-view object correspondence problem as graph clustering that is solved, e.g., by graph cut [20] or k-means [61].

*Spatial correspondence identification* uses spatial relationships of objects to identify their correspondences. For example, iterative closest points (ICP) is a technique commonly used to associate dense points [54]. Correspondence identification is also formulated as a linear assignment problem solved by the Hungarian [1] or Sinkhorn algorithm [59], and a quadratic assignment problem that considers distances between the objects [14, 40]. Recently, angular relationships among the objects are also used for correspondence identification, with a formulation of hypergraph matching that is solved by reweighted random walk [12, 38], tensor block coordinate ascent [49], and Monte Carlo Markov Chain (MCMC) [55].

The appearance and synchronization-based methods require that the object appearance in multiple views must be unique, which cannot well address the scenarios when multiple objects look identical or the same object looks different from different views. Furthermore, most spatial-based methods assume that non-covisible objects only exist in one of the multiple views, but in not all views. Finally, existing methods cannot address perception uncertainty together with non-covisible objects for correspondence identification.

### B. Uncertainty in Perception

Uncertainty in robot perception is traditionally computed as the variance of probability distributions [56]. A widely applied method to address uncertainty is to compensate it by designing an uncertainty model, e.g., applying a sensory uncertainty field for multi-camera tracking [5], modeling odometry uncertainty with fuzzy set for robot motion estimation [10], adding uncertainty to robot joint positions to improve reliability estimation [11], describing uncertainty in point clouds by Gaussian Mixture Model (GMM) to compensate distortion [28, 41], applying a multi-variate Gaussian distribution to model human joints to improve pose prediction [25], and modeling uncertainty of point cloud positions by regression for robot inspection [27]. Such traditional uncertainty models are generally designed for specific robotics tasks.

Recently, Bayesian neural network (BNN) is widely adopted to perform machine perception and quantify its uncertainty, in a variety of applications including monocular depth estimation and segmentation [33], camera localization [3, 33], and object classification [34]. Uncertainty in machine perception based on BNN is generally divided in two categories [16, 31]: epistemic uncertainty and aleatoric uncertainty. *Epistemic uncertainty* is defined as the ambiguity in the BNN learning model e.g., the learning model cannot explain all training data. The epistemic uncertainty can be calculated as the variance of the posterior distribution of BNN model parameters. *Aleatoric uncertainty* is defined as the ambiguity in training data [31] (e.g. caused by over-exposed regions in images when performing monocular depth estimation). The aleatoric uncertainty is computed as the variance of the likelihood distribution of training data, which can be obtained by sampling techniques, such as Monte Carlo sampling [32], or variational inference [42].

We follow the same definitions of perception uncertainties, and, for the first time, address them along with non-covisible objects in a principled framework in order to enable correspondence identification for multi-robot collaborative perception.

### III. THE PROPOSED APPROACH

**Notation.** We write matrices using boldface capital letters, e.g.,  $\mathbf{M} = \{\mathbf{M}_{i,j}\} \in \mathbb{R}^{n \times m}$  with  $\mathbf{M}_{i,j}$  denoting the element in the  $i$ -th row and  $j$ -th column of  $\mathbf{M}$ . Similarly, we also utilize boldface capital letters to represent tensors (i.e., 3D matrices), i.e.,  $\mathbf{T} = \{\mathbf{T}_{i,j,k}\} \in \mathbb{R}^{n \times m \times l}$ . Vectors are written as boldface lowercase letters  $\mathbf{v} \in \mathbb{R}^n$ . In addition, the vectorized form of a matrix  $\mathbf{M} \in \mathbb{R}^{n \times m}$  is represented as  $\mathbf{m} \in \mathbb{R}^{nm}$ , which is a concatenation of each column in  $\mathbf{M}$  into a vector.

#### A. Problem Formulation

We introduce a graph-based representation to address correspondence identification. Given an observation of the environment sensed by a robot, we represent it using an undirected graph  $\mathcal{G} = \{\mathcal{P}, \mathcal{A}, \mathcal{S}\}$ . The node set  $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$  represents the positions of the objects, where  $\mathbf{p}_i = \{x, y, z\}$  denotes the 3D position of the  $i$ -th object and  $n$  is the number of objects observed by the robot.  $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$  denotes the set of attributes to encode appearance and semantic characteristics of the objects, where  $\mathbf{a}_i$  is a vector of attributes of the  $i$ -th object located at  $\mathbf{p}_i$ .  $\mathcal{S} = \{\mathcal{S}^d, \mathcal{S}^a\}$  denotes the spatial relationships among the objects.  $\mathcal{S}^d = \{s_{i,j}^d\}$  denotes the set of distance relationships between a pair of nodes, where  $s_{i,j}^d, i, j = 1, 2, \dots, n, i \neq j$  denotes the distance between  $\mathbf{p}_i$  and  $\mathbf{p}_j$ .  $\mathcal{S}^a = \{s_{i,j,k}^a\}$  denotes the set of angular relationships, where  $s_{i,j,k}^a = [\theta_i, \theta_j, \theta_k], i, j, k = 1, 2, \dots, n, i \neq j \neq k$  is the angles of the triangle constructed by node  $\mathbf{p}_i, \mathbf{p}_j$  and  $\mathbf{p}_k$ . We consider the angular relationship as it is invariant to scale variations and more robust to spatial deformations compared to the distance relationship [18].

In collaborative perception, the objects observed by a pair of robots in their own fields of view can be respectively represented with two graphs  $\mathcal{G} = \{\mathcal{P}, \mathcal{A}, \mathcal{S}\}$  and  $\mathcal{G}' = \{\mathcal{P}', \mathcal{A}', \mathcal{S}'\}$ . Given the graph representations, we can compute the similarity

of the objects' appearance and spatial relationships to facilitate correspondence identification.

- The attribute similarity is computed by

$$\mathbf{A}_{i,i'} = \frac{\mathbf{a}_i \cdot \mathbf{a}'_{i'}}{\|\mathbf{a}_i\| \|\mathbf{a}'_{i'}\|} \quad (1)$$

where  $\mathbf{A}_{i,i'}$  denotes the similarity between attribute vectors  $\mathbf{a}_i \in \mathcal{A}$  and  $\mathbf{a}'_{i'} \in \mathcal{A}'$ . The attribute similarities of all objects represented by the two graphs can be denoted as a matrix  $\mathbf{A} = \{\mathbf{A}_{i,i'}\} \in \mathbb{R}^{n \times n'}$ , as shown in Figure 2(a).

- The distance similarity between two pairs of objects can be calculated by

$$\mathbf{D}_{ii',jj'} = \exp\left(-\frac{1}{\gamma}(s_{i,j}^d - s_{i',j'}^d)^2\right) \quad (2)$$

where  $\mathbf{D}_{ii',jj'}$  is the similarity between distance  $s_{i,j}^d \in \mathcal{S}^d$  and distance  $s_{i',j'}^d \in \mathcal{S}^d$ . We use an exponential function parameterized by  $\gamma$  to normalize  $\mathbf{D}_{ii',jj'} \in (0, 1]$ . The distance similarities of all pairs of objects represented by two graphs are denoted by the matrix  $\mathbf{D} = \{\mathbf{D}_{ii',jj'}\} \in \mathbb{R}^{nn' \times nn'}$ , as shown in Figure 2(b).

- The angular similarity between two triangles constructed by three nodes in each graph is defined as follows:

$$\mathbf{T}_{ii',jj',kk'} = \exp\left(-\frac{1}{\gamma} \sum_{p \in \{i,j,k; i',j',k'\}} |\cos(\theta_p) - \cos(\theta'_q)|\right) \quad (3)$$

where  $\mathbf{T}_{ii',jj',kk'}$  denotes the similarity between triangle  $s_{i,j,k}^a \in \mathcal{S}^a$  and triangle  $s_{i',j',k'}^a \in \mathcal{S}^a$ . The angular similarities of all objects encoded by the two graphs are denoted by the tensor  $\mathbf{T} = \{\mathbf{T}_{ii',jj',kk'}\} \in \mathbb{R}^{nn' \times nn' \times nn'}$ , as shown in Figure 2(c).

Then, we formulate correspondence identification as a graph matching problem that integrates the similarities of the object attributes and spatial relationships into a unified optimization framework to identify correspondence of the objects observed by a pair of robots in collaborative perception. Mathematically, the problem formulation can be expressed as follows:

$$\begin{aligned} \max_{\mathbf{X}} \quad & \mathbf{A}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{D} \mathbf{x} + \mathbf{T} \otimes_1 \mathbf{x} \otimes_2 \mathbf{x} \otimes_3 \mathbf{x} \\ \text{s.t.} \quad & \mathbf{X} \mathbf{1}_{n' \times 1} \leq \mathbf{1}_{n \times 1}, \mathbf{X}^\top \mathbf{1}_{n \times 1} \leq \mathbf{1}_{n' \times 1} \end{aligned} \quad (4)$$

where  $\mathbf{X} \in \mathbb{R}^{n \times n'}$  is the correspondence matrix,  $\mathbf{x} = \{x_{ii'}\} \in \{0, 1\}^{nn'}$  is the vectorized form of  $\mathbf{X}$ , with  $x_{ii'} = 1$  indicating that the  $i$ -th object in  $\mathcal{G}$  corresponds to the  $i'$ -th object in  $\mathcal{G}'$  (otherwise  $x_{ii'} = 0$ ),  $\otimes$  denotes tensor product,  $\otimes_l, l = 1, 2, 3$  denotes multiplication between  $\mathbf{x}$  and the mode- $l$  matricization of  $\mathbf{T}$  [37], and  $\mathbf{1}$  denotes a vector with all ones.

The objective function in Eq. (4) denotes the overall similarity, given the correspondence matrix  $\mathbf{X}$ . The first term denotes the accumulated attribute similarity, the second term denotes the accumulated distance similarity, and the third term denotes the accumulated angular similarity of the objects encoded by the pair of graphs. The constraints in Eq. (4) are introduced to enforce each row and column in  $\mathbf{X}$  to at most have one element equal to 1, thus guaranteeing one-to-one correspondences.

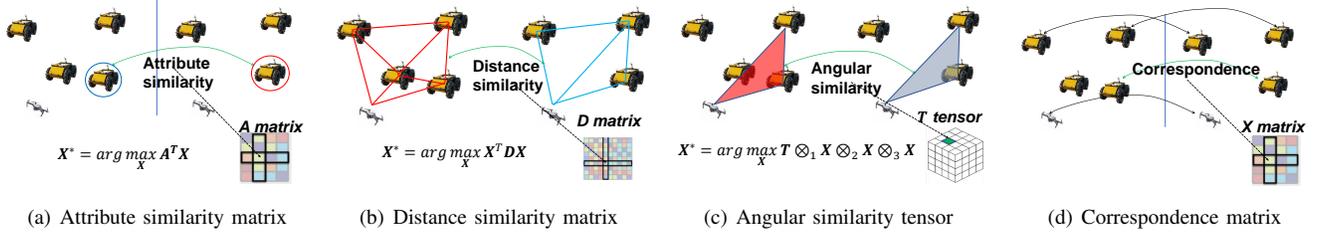


Fig. 2. Illustrations of the defined attribute similarity matrix  $\mathbf{A}$ , distance similarity matrix  $\mathbf{D}$ , angular similarity tensor  $\mathbf{T}$ , and correspondence matrix  $\mathbf{X}$ , given two graphs that represent objects (e.g., UAVs and UGVs) observed by a pair of robots.

### B. Addressing Uncertainty and Non-Covisibility

Based on our problem formulation in Eq. (4) that formulates correspondence identification as an optimization problem, we propose a novel solution by designing new regularization terms to regularize the optimization in order to explicitly address the challenges of uncertainty and non-covisibility, which have not been well addressed for correspondence identification.

1) *Addressing uncertainty*: Mainly due to sensor resolution limit and noise, measurement scenario variations, and perception model bias, uncertainty always exists in robot perception.

The *uncertainty in object attributes* is defined as the average of the variances of individual elements in the attribute vector. For example, when using semantic attributes, which are one of the most widely used object attributes [23, 26], we can utilize Bayesian neural networks (BNN) [31] to directly estimate the attribute uncertainty. By sampling the posterior distribution of the BNN model parameters, we can compute the variance of the model parameters as the epistemic uncertainty that captures the ambiguity in the BNN model. By sampling the likelihood distribution of the predicted semantic labels, we can calculate the variance of the semantic labels as the aleatoric uncertainty that captures the ambiguity in data. Then, the uncertainty of the semantic attributes can be computed as a sum of the epistemic and aleatoric uncertainties.

Formally, we use  $v_i$  to denote the uncertainty of the attribute vector  $\mathbf{a}_i$  of the  $i$ -th object computed as the average of the variances of individual elements in  $\mathbf{a}_i$ , and  $\mathbf{v} = [v_1, v_2, \dots, v_n]$  to represent the attribute uncertainties of all  $n$  objects encoded by graph  $\mathcal{G}$ . Given  $\mathbf{v}$  and  $\mathbf{v}'$  from  $\mathcal{G}$  and  $\mathcal{G}'$  respectively, the attribute uncertainty matrix  $\mathbf{V}$  is calculated as follows:

$$\mathbf{V} = \mathbf{v} \oplus \mathbf{v}'^T \quad (5)$$

where  $\oplus$  denotes the kronecker plus [47], and  $\mathbf{V} = \{\mathbf{V}_{i,i'}\} \in \mathbb{R}^{n \times n'}$  is the attribute uncertainty matrix, with  $\mathbf{V}_{i,i'} = v_i + v_{i'}$  indicating the uncertainty of using  $\mathbf{a}_i$  and  $\mathbf{a}_{i'}$  to compute the attribute similarity  $\mathbf{A}_{i,i'}$  in Eq. (1). We consider  $\mathbf{A}_{i,i'}$  to provide more important information if its uncertainty  $\mathbf{V}_{i,i'}$  has a smaller value. Accordingly, we compute a weight matrix  $\mathbf{W}^a$  for the attribute similarity matrix  $\mathbf{A}$  based on  $\mathbf{V}$ :

$$\mathbf{W}^a = \exp\left(-\frac{1}{\sigma} \mathbf{V}\right) \quad (6)$$

where  $\sigma$  denotes the parameter of the normalization function, and  $\mathbf{W}^a = \{\mathbf{W}_{i,i'}^a\} \in \mathbb{R}^{n \times n'}$  is the weight matrix with  $\mathbf{W}_{i,i'}^a$  indicating the importance (in terms of certainty) of  $\mathbf{A}_{i,i'}$ .

The *uncertainty in object positions* is defined as the average of the variances of pixel-level depth values of an object. When only monocular visual observations are available, BNNs can be used to estimate the depth values, which are able to directly provide the uncertainty. Similarly, this uncertainty is a sum of the epistemic uncertainty in the BNN model and the aleatoric uncertainty in the data. Mathematically, we use  $u_i$  to denote the position uncertainty of the  $i$ -th object, which is calculated as the average of the variances of the object's depth values, and  $\mathbf{u} = [u_1, u_2, \dots, u_n]$  to represent the position uncertainties of all objects by  $\mathcal{G}$ . Given  $\mathbf{u}$  and  $\mathbf{u}'$  from  $\mathcal{G}$  and  $\mathcal{G}'$  respectively, the position uncertainty matrix  $\mathbf{U}$  is computed by:

$$\mathbf{U} = \mathbf{u} \oplus \mathbf{u}'^T \quad (7)$$

where  $\mathbf{U} = \{\mathbf{U}_{i,i'}\} \in \mathbb{R}^{n \times n'}$  with  $\mathbf{U}_{i,i'} = u_i + u_{i'}$  indicating the position uncertainty of a pair of objects.

According to Eq. (2), we can compute the similarity  $\mathbf{D}_{ii',jj'}$  between distance  $s_{i,j}^d \in \mathcal{S}^d$  and distance  $s_{i',j'}^d \in \mathcal{S}^d$  based on the object positions, and similarly, we assume that  $\mathbf{D}_{ii',jj'}$  has a larger weight if it has a lower uncertainty. This weight matrix  $\mathbf{W}^d$  of the distance similarity matrix  $\mathbf{D}$  can be calculated as:

$$\mathbf{W}^d = \exp\left(-\frac{1}{\sigma} (\mathbf{U} \oplus \mathbf{U}^T)\right) \quad (8)$$

where  $\mathbf{W}^d = \{\mathbf{W}_{ii',jj'}^d\} \in \mathbb{R}^{nn' \times nn'}$ , and  $\mathbf{W}_{ii',jj'}^d$  represents the importance of  $\mathbf{D}_{ii',jj'}$  and is computed based on two pairs of objects. Similarly, we compute the importance tensor  $\mathbf{W}^t$  of the angular similarity tensor  $\mathbf{T}$  in Eq. (3) as follows:

$$\mathbf{W}^t = \exp\left(-\frac{1}{\sigma} (\mathbf{U} \oplus (\mathbf{U} \oplus \mathbf{U}^T)^T)\right) \quad (9)$$

where  $\mathbf{W}^t = \{\mathbf{W}_{ii',jj',kk'}^t\} \in \mathbb{R}^{nn' \times nn' \times nn'}$ , and  $\mathbf{W}_{ii',jj',kk'}^t$  is the weight (in terms of certainty) of the angular similarity  $\mathbf{T}_{ii',jj',kk'}$  between triangles  $s_{i,j,k}^a \in \mathcal{S}^a$  and  $s_{i',j',k'}^a \in \mathcal{S}^a$ .

Then, the weights  $\mathbf{W}^a$ ,  $\mathbf{W}^d$ , and  $\mathbf{W}^t$  are utilized to encode the importance of the similarities in  $\mathbf{A}$ ,  $\mathbf{D}$ , and  $\mathbf{T}$ , in order to make the method to rely more on the similarities with a lower uncertainty for correspondence identification:

$$\begin{aligned} \max_{\mathbf{X}} (\mathbf{A} \circ \mathbf{W}^a)^T \mathbf{x} + \mathbf{x}^T (\mathbf{D} \circ \mathbf{W}^d) \mathbf{x} + \mathbf{T} \circ \mathbf{W}^t \otimes_1 \mathbf{x} \otimes_2 \mathbf{x} \otimes_3 \mathbf{x} \\ \text{s.t. } \mathbf{X} \mathbf{1}_{n' \times 1} \leq \mathbf{1}_{n \times 1}, \mathbf{X}^T \mathbf{1}_{n \times 1} \leq \mathbf{1}_{n' \times 1} \end{aligned} \quad (10)$$

In addition, we introduce a new regularization term over the correspondence matrix  $\mathbf{X}$  to control the sum of uncertainties.

Intuitively, if two objects have larger uncertainties, it may still be inappropriate to match them, even though they have a large similarity score. This regularization term can be integrated into our problem formulation in the unified regularized optimization framework:

$$\begin{aligned} \max_{\mathbf{X}} & (\mathbf{A} \circ \mathbf{W}^a)^\top \mathbf{x} + \mathbf{x}^\top (\mathbf{D} \circ \mathbf{W}^d) \mathbf{x} + \mathbf{T} \circ \mathbf{W}^t \otimes_1 \mathbf{x} \otimes_2 \mathbf{x} \otimes_3 \mathbf{x} \\ & - \lambda_1 \|(\mathbf{V} + \mathbf{U}) \circ \mathbf{X}\|^2 \\ \text{s.t.} & \quad \mathbf{X} \mathbf{1}_{n' \times 1} \leq \mathbf{1}_{n \times 1}, \mathbf{X}^\top \mathbf{1}_{n \times 1} \leq \mathbf{1}_{n' \times 1} \end{aligned} \quad (11)$$

where  $\circ$  is entry-wise product. The regularization term  $\|(\mathbf{V} + \mathbf{U}) \circ \mathbf{X}\|^2$  denotes the overall attribute and position uncertainty given  $\mathbf{X}$ . The hyper-parameter  $\lambda_1$  is introduced to balance the maximization of the overall similarity and the minimization of the overall uncertainty.

2) *Address non-covisibility*: Non-covisible objects usually significantly increase the number of incorrect correspondences, because an object observed by one robot may not be observed by other robots (e.g., due to limited field of view or occlusion), and thus correspondences may not exist. In order to explicitly address this issue, we introduce the regularization term  $\|\mathbf{X}\|^2$  to reduce the number of correspondences:

$$\begin{aligned} \max_{\mathbf{X}} & (\mathbf{A} \circ \mathbf{W}^a)^\top \mathbf{x} + \mathbf{x}^\top (\mathbf{D} \circ \mathbf{W}^d) \mathbf{x} + \mathbf{T} \circ \mathbf{W}^t \otimes_1 \mathbf{x} \otimes_2 \mathbf{x} \otimes_3 \mathbf{x} \\ & - \lambda_2 \|\mathbf{X}\|^2 \\ \text{s.t.} & \quad \mathbf{X} \mathbf{1}_{n' \times 1} \leq \mathbf{1}_{n \times 1}, \mathbf{X}^\top \mathbf{1}_{n \times 1} \leq \mathbf{1}_{n' \times 1} \end{aligned} \quad (12)$$

where  $\lambda_2$  is the hyper-parameter to balance the overall similarity and the regularization term. When the number of identified object correspondences increases, both the value of  $\|\mathbf{X}\|^2$  and the overall similarity increase. One correspondence is added to  $\mathbf{X}$  only if the increase of the overall similarity caused by the newly added correspondence is larger than the penalty resulted from the regularization. Accordingly, correspondences among non-covisible objects often having smaller similarities are less likely to be added to  $\mathbf{X}$ , and co-visible objects that have larger similarities are more likely to be added to  $\mathbf{X}$  and identified.

In summary, after integrating both regularization terms into the unified mathematical framework of regularized constrained optimization, our final graph matching formulation to address correspondence identification with non-covisible objects under uncertainty becomes:

$$\begin{aligned} \max_{\mathbf{X}} & (\mathbf{A} \circ \mathbf{W}^a)^\top \mathbf{x} + \mathbf{x}^\top (\mathbf{D} \circ \mathbf{W}^d) \mathbf{x} + \mathbf{T} \circ \mathbf{W}^t \otimes_1 \mathbf{x} \otimes_2 \mathbf{x} \otimes_3 \mathbf{x} \\ & - \lambda_1 \|(\mathbf{V} + \mathbf{U}) \circ \mathbf{X}\|^2 - \lambda_2 \|\mathbf{X}\|^2 \\ \text{s.t.} & \quad \mathbf{X} \mathbf{1}_{n' \times 1} \leq \mathbf{1}_{n \times 1}, \mathbf{X}^\top \mathbf{1}_{n \times 1} \leq \mathbf{1}_{n' \times 1} \end{aligned} \quad (13)$$

### C. Optimization Algorithm

Since the proposed constrained optimization formulation in Eq. (13) is a non-convex problem and has regularization terms, the commonly used optimization methods for graph matching, e.g., based upon reweighted random walks [12, 38], cannot be directly utilized to solve the problem. Thus, we design a new heuristic optimization algorithm based on Markov chain Monte Carlo (MCMC) sampling [55].

---

**Algorithm 1:** The proposed solver to solve the formulated non-convex regularized constrained optimization problem in Eq. (13).

---

**Input** :  $\mathbf{T} \in \mathbb{R}^{nn' \times nn' \times nn'}$ ,  $\mathbf{D} \in \mathbb{R}^{nn' \times nn'}$ , and  $\mathbf{A}, \mathbf{U}, \mathbf{V} \in \mathbb{R}^{n \times n'}$   
**Output** :  $\mathbf{x} \in \{0, 1\}^{nn'}$

- 1: Initialize  $\mathbf{y} \in \{0, 1\}^n$  (belongs to  $\mathcal{G}$ ) and  $T$ ;
- 2: **while**  $T > T_f$  **do**
- 3:     Compute the active node distribution  $q(\mathbf{y} \rightarrow \hat{\mathbf{y}})$  in Eq. (16);
- 4:     Sample  $\mathbf{y}$  from  $q$ ;
- 5:     Calculate the acceptance ratio  $\alpha(\mathbf{y} \rightarrow \hat{\mathbf{y}})$  in Eq. (17);
- 6:     **if**  $\alpha(\mathbf{y} \rightarrow \hat{\mathbf{y}}) > \epsilon$  **then**
- 7:          $\mathbf{y} \rightarrow \hat{\mathbf{y}}$ ;
- 8:         **if**  $P(\hat{\mathbf{y}}) > P(\mathbf{y})$  in Eq. (14) **then**
- 9:             Calculate  $\mathbf{x}^*$  in Eq. (15);
- 10:             Update state  $\mathbf{y} = \hat{\mathbf{y}}$ ;
- 11:         **end**
- 12:     **end**
- 13:      $T = \xi T$ ;
- 14: **end**
- 15: **return**  $\mathbf{x}^*$

---

We construct a Markov chain on the state space  $\mathcal{Y} = \{\mathbf{y} | \mathbf{y} \in \{0, 1\}^n\}$ , whose stationary distribution describes the matching objects, and  $\mathbf{y}$  denotes a subset of the objects in graph  $\mathcal{G}$ ; if  $\mathbf{y}_i = 1$ , we set the  $i$ -th object as active and use it for graph matching. Then, we convert the problem in Eq. (13) to the following:

$$\begin{aligned} P(\mathbf{y}) = & \exp((\mathbf{A} \circ \mathbf{W}^a)^\top \pi(\mathbf{y}) + \pi(\mathbf{y})^\top (\mathbf{D} \circ \mathbf{W}^d) \pi(\mathbf{y}) \\ & + \mathbf{T} \circ \mathbf{W}^t \otimes_1 \pi(\mathbf{y}) \otimes_2 \pi(\mathbf{y}) \otimes_3 \pi(\mathbf{y}) \\ & - \lambda_1 \|(\mathbf{U} + \mathbf{V}) \circ \pi(\mathbf{y})\|^2 - \lambda_2 \|\pi(\mathbf{y})\|^2 \end{aligned} \quad (14)$$

where  $\pi(\mathbf{y}) \in \mathbb{R}^{nn'}$  denotes the correspondences given active nodes on  $\mathcal{Y}$ , and  $P(\mathbf{y})$  denotes the overall similarity given the correspondences. Formally,  $\pi(\mathbf{y})$  is computed as follows:

$$\begin{aligned} \pi(\mathbf{y}) = & \max_{\mathbf{X}} (\mathbf{A} \circ \mathbf{W}^a)^\top \mathbf{x} + \mathbf{x}^\top (\mathbf{D} \circ \mathbf{W}^d) \mathbf{x} \\ & + \mathbf{T} \circ \mathbf{W}^t \otimes_1 \mathbf{x} \otimes_2 \mathbf{x} \otimes_3 \mathbf{x} \\ \text{s.t.} & \quad \sum_i \mathbf{X}_{ij} = 1, \text{ if } \mathbf{y}_i = 1 \\ & \quad \sum_j \mathbf{X}_{ij} = 0, \text{ if } \mathbf{y}_i = 0 \\ & \quad \mathbf{X} \mathbf{1}_{n' \times 1} \leq \mathbf{1}_{n \times 1}, \mathbf{X}^\top \mathbf{1}_{n \times 1} \leq \mathbf{1}_{n' \times 1} \end{aligned} \quad (15)$$

Given the active nodes encoded by  $\mathbf{y}$ , the first two constraints restrict  $\mathbf{X}$  to only include correspondences by the active nodes, and the final correspondences is obtained from solving  $\pi(\mathbf{y})$ .

In order to select optimal active nodes encoded by state  $\mathbf{y}$ , we design the rule to iteratively update  $\mathbf{y}$ :

$$q(\mathbf{y}, \hat{\mathbf{y}}) = \begin{cases} \alpha \exp\left(-\frac{1}{\gamma} (\hat{\mathbf{y}}^\top (\mathbf{v} + \mathbf{u}))\right) & \|\hat{\mathbf{y}}\|_1 = \|\mathbf{y}\|_1 + 1 \\ (1 - \alpha) \frac{1}{\|\mathbf{y}\|_1} & \|\hat{\mathbf{y}}\|_1 = \|\mathbf{y}\|_1 - 1 \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

where  $q(\mathbf{y}, \hat{\mathbf{y}})$  denotes the transition distribution to update  $\mathbf{y}$  to  $\hat{\mathbf{y}}$ , and  $\hat{\mathbf{y}}$  is obtained by sampling from the distribution  $q(\mathbf{y}, \hat{\mathbf{y}})$ . There are two modes to update  $\hat{\mathbf{y}}$  based on  $q(\mathbf{y}, \hat{\mathbf{y}})$ , including adding one active node  $\|\hat{\mathbf{y}}\|_1 = \|\mathbf{y}\|_1 + 1$ , and deleting one active node  $\|\hat{\mathbf{y}}\|_1 = \|\mathbf{y}\|_1 - 1$ ; otherwise the probability equals to 0. The mode selection is controlled by  $\alpha$ , which is set to 0.5, meaning that there is a 50% probability to add or delete one active node from  $\mathbf{y}$  in each update.

When adding an active node, we select nodes with a small uncertainty in order to use low-uncertainty nodes to compute the correspondence. Accordingly, the attribute uncertainty  $\mathbf{v}$  and position uncertainty  $\mathbf{u}$  are used to compute the probability of adding an active node when updating  $\hat{\mathbf{y}}$ . If the uncertainties are high, the probability of updating to  $\hat{\mathbf{y}}$  is low. The probability of deleting a node from  $\mathbf{y}$  follows a uniform distribution decided by the number of active nodes, meaning that all active nodes are treated equally during deletion.

In addition, to ensure that the Markov chain converges to a stable distribution, the state update in Markov chain should be subject to the detailed balance condition [9], which means the the designed Markov chain is reversible. According to [55], the acceptance ratio of state update is designed as:

$$\alpha(\mathbf{y}, \hat{\mathbf{y}}) = \min \left( \frac{P(\hat{\mathbf{y}})q(\hat{\mathbf{y}}, \mathbf{y})}{P(\mathbf{y})q(\mathbf{y}, \hat{\mathbf{y}})}, 1 \right) \quad (17)$$

The proposed algorithm is shown in Algorithm 1. The end condition is controlled by the annealing temperature  $T$ , and the algorithm stops if  $T$  reduces to a predefined value  $T_f$  with the annealing rate  $\xi$ .

**Complexity.** The complexity to solve the optimization problem in Eq. (11) is  $O(n^6)$ , which is dominated by  $\mathbf{T}$ . When we apply a nearest neighborhood search to compute local matches, the complexity reduces to  $O(n^2k)$ , where  $k$  is the number of nearest neighborhoods. In this paper, we set  $k = n^2$  and reduce the complexity to  $O(n^4)$ .

## IV. EXPERIMENTAL RESULTS

### A. Experiment Setup

We utilize both robotics simulations and physical robots to evaluate our method for correspondence identification in multi-robot collaborative perception in the three scenarios, including simulated connected autonomous driving (CAD), simulated multi-robot coordination (S-MRC), and real-world multi-robot coordination (R-MRC)<sup>1</sup>. Each of the datasets includes 50 pairs of video instances with each video lasting around 5 seconds. Each video instance includes a pair of monocular RGB images observed by two robots from different viewpoints, as well as the ground truth of object correspondence that is obtained from the simulations (CAD and S-MRC) or the QR code (R-MRC). We use the attribute features from [35], and their uncertainties are obtained from BNN models [31], as shown in Figure 3(b). Object positions are computed from depth estimation [31] and the position uncertainties are also obtained from BNN models, which are shown in Figures 3(c) and 3(d), respectively.

We adopt precision and recall as metrics to evaluate the performance of correspondence identification, following [55, 63]. Given the correspondences of covisible objects, precision is defined as the ratio of correspondences of co-visible objects over all retrieved correspondences, and recall is defined as the ratio of the retrieved correspondences of co-visible objects over all ground truth correspondences of co-visible objects.

We implement the full version of our approach that includes both regularization terms to explicitly address uncertainty and non-covisibility with hyper-parameters  $\lambda_1 = 0.1$  and  $\lambda_2 = 0.4$ . They are decided using sensitive analysis in our experiments. Intuitively, non-visibility is a more severe challenge for correspondence identification, as non-visibility results in missing data. Uncertainty is mainly caused by noise in the input data, and is less severe than non-visibility. This explains why  $\lambda_2$  is greater than  $\lambda_1$  in general. In addition, we implement two baseline methods by setting  $\lambda_1 = 0$  that only uses the non-covisibility regularization without considering uncertainty, and by setting  $\lambda_2 = 0$  that only uses the uncertainty regularization without considering non-covisibility.

Also, for comparison, we implement six previous correspondence identification techniques based on object appearance or spatial information. In terms of the spatial-based methods, we implement (1) pairwise graph matching **RRWM** [14] that uses the distance similarity to identify correspondences, (2) iterative closest point **ICP** [4] that iteratively minimizes the distances of two graphs, two hypergraph matching techniques, including (3) **BCAGM** [49] and (4) **RRWHM** [38] that use angular spatial relationships of the objects to identify correspondences. To compare with the attribute-based methods, (5) we implement a CNN-based attribute learning and matching approach for object re-identification (**ReId**) [65], and (6) an approach based upon multi-order similarities (**MOS**) [12] that consider both appearance and spatial relationships.

### B. Results on Connected Autonomous Driving Simulations

Various street objects are contained in the CAD simulation, including different vehicles, pedestrians, traffic lights, and road signs. The views from both connected vehicles contain strong occlusions and large numbers of non-covisible objects.

The quantitative results are presented in Table I. We observe that our complete approach outperforms two baseline methods, indicating the benefit of addressing both perception uncertainty and non-covisibility. Comparisons with the previous methods are also presented in Table I. It is observed that distance-based techniques (ICP and RRWM) perform badly, because of spatial deformations of the spatial distances resulted from the position uncertainty. The angular-based techniques (BCAGM and RRWHM) achieve improved performance, since the triangle relationships of objects are more robust to perspective changes. ReId improves the performance by using attribute features of objects, which is more robust to appearance variations. MOS further improves the performance through integrating visual and spatial information of objects. Our approach significantly outperforms the previous methods because of its capability of

<sup>1</sup>The datasets are available at: <http://hcr.mines.edu/project/civr.html>.

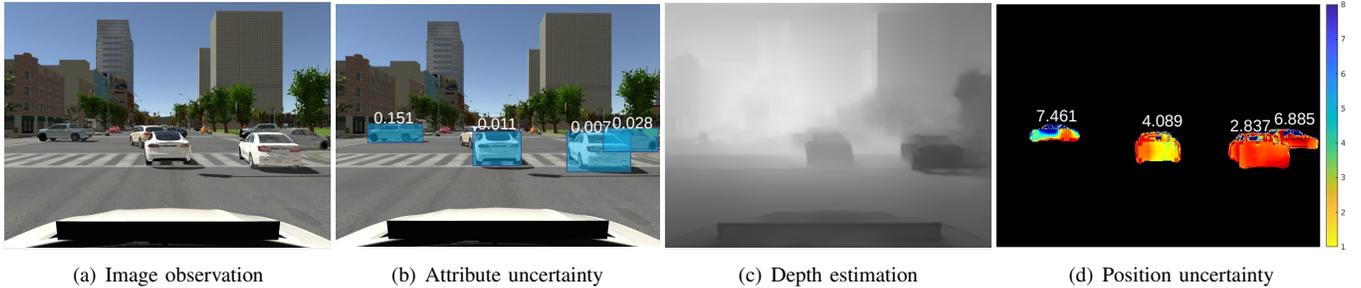


Fig. 3. Illustration of the image observed by a vehicle in CAD, attribute uncertainty, depth estimation, and position uncertainty. The numbers in Figures 3(b) and 3(d) denote average attribute and position uncertainties for each object, respectively. In Figure 3(d), a warmer color within an object region indicates a lower uncertainty.

integrating both attribute and spatial relationships, as well as addressing uncertainty and non-covisible objects.

The qualitative results of our approach in CAD are presented in Figure 4(c), which demonstrates that our approach correctly identifies correspondences of street objects. For comparison, we also include the qualitative results obtained by ReId and MOS in Figure 4(a) and Figure 4(b), respectively. Since most objects in this situation have similar appearance and attributes (e.g., many white and gray vehicles), ReId cannot well identify the object correspondences. Although MOS can identify most correspondences of the objects, but the precision is low. The reason is because that MOS always maximizes the number of correspondences to obtain the highest similarity value, without considering the non-covisible objects that cannot be matched.

In order to further evaluate the robustness of our approach, we manually increase perception uncertainty in attributes and positions of the objects, and then evaluate the result variations. Specifically, given a uncertainty rate (i.e., 0-15%), we set the uncertainty value to  $uncertainty \times (1 + uncertainty\_rate)$ . The performance variations on precision and recall with respect to different attribute and position uncertainty rate are demonstrated in Figure 5. It is observed that the performance of our proposed approach gradually decreases with small fluctuations as the increase of the uncertainty rate in attributes and positions. We also observe that our method obtains robust performance with the uncertainty rate within 10%.

When running our approach on a Linux machine with an i7 3.0GHz CPU, 16G memory and no GPU, the execution speed is around 5Hz, if  $n = 15$  and 200 samplings are used. When parallel MCMC-sampling [46] is applied, the execution speed can be further improved to around 40Hz on an 8-cores CPU.

### C. Results on Multi-robot Coordination Simulations

We evaluate our approach in multi-robot coordination simulations. The object instances used in the simulations include a team of Husky UGVs with identical appearances. The objects are observed by another two robots with partially overlapped views. This simulator is implemented by integrating Unity for visualization with ROS for robot perception and control.

The qualitative results in S-MRC are shown in Figure 4(f). It demonstrates that our approach is able to correctly identifies correspondences of the UGVs from two views. Comparisons

TABLE I  
QUANTITATIVE RESULTS OF OUR APPROACH, AND COMPARISONS WITH PREVIOUS AND BASELINE METHODS IN CAD, S-MRC AND R-MRC.

Method	CAD		S-MRC		R-MRC	
	Precision	Recall	Precision	Recall	Precision	Recall
RRWM [14]	0.032	0.029	0.289	0.291	0.131	0.106
ICP [4]	0.061	0.069	0.302	0.311	0.089	0.116
BCAGM [49]	0.154	0.186	0.503	0.587	0.201	0.256
RRWHM [38]	0.161	0.173	0.524	0.592	0.219	0.221
ReId [65]	0.238	0.264	0.321	0.472	0.396	0.508
MOS [12]	0.571	0.681	0.641	0.658	0.559	0.702
Ours, $\lambda_1 = 0$	0.611	0.641	0.678	0.661	0.563	0.708
Ours, $\lambda_2 = 0$	0.608	0.684	0.685	0.688	0.565	0.711
<b>Ours</b>	<b>0.659</b>	<b>0.718</b>	<b>0.701</b>	<b>0.711</b>	<b>0.575</b>	<b>0.723</b>

with ReId and MOS are also presented in Figure 4. We observe that ReId does not work well since most objects are identical; the identical objects are matched by ReId purely based on their processing order in the approach. The correspondence results obtained by MOS has low precision because of the uncertainty in positions. Our method correctly identifies covisible objects' correspondences under uncertainty in the simulations.

The quantitative results in S-MRC are presented in Table I. The table shows that the baseline method using the uncertainty regularization without considering non-covisibility ( $\lambda_2 = 0$ ) obtains better performance compared with the baseline method without applying the uncertainty regularization ( $\lambda_1 = 0$ ) in this scenario. This is because uncertainty is the main challenge in S-MRC, due to the low resolution of observed images, the low texture of the objects, and their long distance for the cameras. By using both regularization terms, our full approach can still improve performance. Quantitative comparisons with previous techniques are also presented in Table I. It is observed that our proposed approach outperforms the previous methods on both precision and recall. The results demonstrate that, because the S-MRC scenarios contain significant uncertainty, explicitly addressing the uncertainties in attributes and positions of the objects is necessary.

### D. Results on Real-world Multi-robot Coordination

We perform additional evaluation on multi-robot coordination using physical robots. In R-MRC, the object instances are different robots observed from different perspectives (overhead view and side view), in which most of the robots have the same

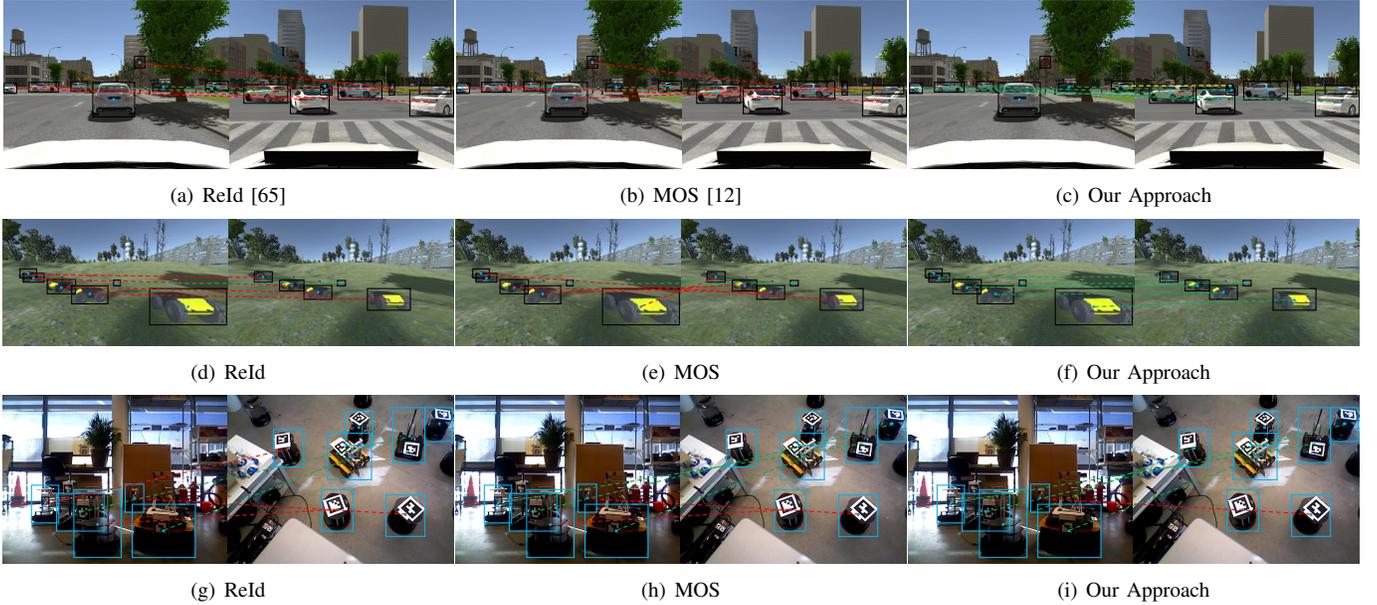


Fig. 4. Qualitative experimental results of our approach over CAD (first row), S-MRC (second row) and R-MRC (third row), and comparisons with the ReId and MOS methods. Green lines denote correct correspondences, red lines denote incorrect correspondences; and yellow lines denote missing correspondences (i.e., false negatives). [Best viewed in color.]

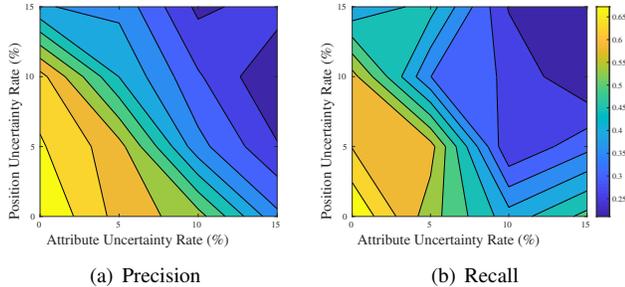


Fig. 5. Robustness analysis of our approach with respect to the attribute and depth uncertainties in the CAD simulations. The uncertainty rate denotes the percentage of additional uncertainties added to object attributes and positions.

type with the identical appearance. The overhead view from a drone can well observe the objects, but the side view obtained by a ground robot has strong occlusions and a smaller field of view. Also, since the objects cannot be well observed from the side view, they are not well perceived, and the estimated robot positions and attributes include significant uncertainties.

The qualitative results in R-MRC are demonstrated in Figure 4. The results indicate that ReId can identify unique objects but can not identify the correspondences of identical objects. MOS obtains an improved performance, but MOS still obtains incorrect correspondences caused by the large uncertainty in the estimated object positions. By addressing both attribute and position uncertainties, our approach obtains the best results on object correspondence identification in these experiments. The quantitative results of correspondence identification in R-MRC are listed in Table I. Since there exist several robots with unique appearances in R-MRC, ReId correctly identifies their correspondences, thus obtaining an improved result compared

to its result in S-MRC. On the other hand, methods (RRWM, ICP, BCAGM, and RRWHM) based upon spatial relationships do not perform well, because of the large uncertainty in object positions. MOS that combines the attribute information with spatial relationships obtains an improved result. By explicitly addressing perception uncertainties and the non-covisible objects, our full approach obtains the best performance in the experiments of multi-robot coordination using physical robots.

## V. CONCLUSION

Correspondence identification is a critical ability for a group of robots to consistently refer to the same objects within their own fields of view. Perception uncertainties and non-covisible objects are two of the biggest challenges to enable this ability. We propose a novel regularized graph matching approach that formulates correspondence identification as an optimization-based graph matching problem with two novel regularization terms to explicitly address uncertainty and non-covisibility. Furthermore, a new sampling-based optimization algorithm is implemented to solve the formulated non-convex regularized constrained optimization problem. Extensive experiments are conducted to evaluate our method both in robotics simulations and using physical robots, in the applications of connected autonomous driving and multi-robot coordination. The experimental results have shown that our approach obtains the state-of-the-art performance for correspondence identification with non-covisible objects under uncertainty.

## ACKNOWLEDGMENTS

This work was supported by NSF CAREER Award IIS-1942056, NSF CNS-1823245, DOT 693JK31850005CAAP, and ARL DCIST CRA W911NF-17-2-0181. The authors also would like to thank the anonymous reviewers and the area chair for their feedback.

## REFERENCES

- [1] HA Almohamad and Salih O Duffuaa. A linear programming approach for the weighted graph matching problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(5):522–525, 1993.
- [2] Rosario Aragues, Eduardo Montijano, and Carlos Sagues. Consistent data association in multi-robot systems with limited communications. In *Robotics: Science and Systems*, 2011.
- [3] Lorenzo Bertoni, Sven Kreiss, and Alexandre Alahi. MonoLoco: Monocular 3D Pedestrian Localization and Uncertainty Estimation. In *IEEE International Conference on Computer Vision*, 2019.
- [4] Paul J Besl and Neil D McKay. Method for registration of 3D shapes. In *Sensor Fusion*, volume 1611, pages 586–606, 1992.
- [5] James Black and Tim Ellis. Multi-camera image measurement and correspondence. *Measurement*, 32(1):61–71, 2002.
- [6] Elizabeth R Boroson and Nora Ayanian. 3D Keypoint Repeatability for Heterogeneous Multi-Robot SLAM. In *IEEE International Conference on Robotics and Automation*, 2019.
- [7] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.
- [8] Manuele Brambilla, Eliseo Ferrante, Mauro Birattari, and Marco Dorigo. Swarm robotics: a review from the swarm engineering perspective. *Swarm Intelligence*, 7(1):1–41, 2013.
- [9] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.
- [10] Pär Buschka, Alessandro Saffiotti, and Zbigniew Wasik. Fuzzy landmark-based localization for a legged robot. In *International Conference on Intelligent Robots and Systems*, volume 2, pages 1205–1210, 2000.
- [11] Carlos Carreras and Ian D Walker. Interval methods for fault-tree analysis in robotics. *IEEE Transactions on Reliability*, 50(1):3–11, 2001.
- [12] Hyung Jin Chang, Tobias Fischer, Maxime Petit, Martina Zambelli, and Yiannis Demiris. Learning kinematic structure correspondences using multi-order similarities. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2920–2934, 2017.
- [13] Yanbei Chen, Xiatian Zhu, and Shaogang Gong. Person re-identification by deep learning multi-scale representations. In *IEEE International Conference on Computer Vision*, 2017.
- [14] Minsu Cho, Jungmin Lee, and Kyoung Mu Lee. Reweighted random walks for graph matching. In *European Conference on Computer Vision*, 2010.
- [15] Soon-Jo Chung, Aditya Avinash Paranjape, Philip Dames, Shaojie Shen, and Vijay Kumar. A survey on aerial swarm robotics. *IEEE Transactions on Robotics*, 34(4):837–855, 2018.
- [16] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? Does it matter? *Structural Safety*, 31(2):105–112, 2009.
- [17] Mehmet Dogar, Andrew Spielberg, Stuart Baker, and Daniela Rus. Multi-robot grasp planning for sequential assembly operations. *Autonomous Robots*, 43(3):649–664, 2019.
- [18] Olivier Duchenne, Francis Bach, In-So Kweon, and Jean Ponce. A tensor-based algorithm for high-order graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2383–2395, 2011.
- [19] Jakob Engel, Thomas Schöps, and Daniel Cremers. LSD-SLAM: Large-scale direct monocular SLAM. In *European Conference on Computer Vision*, 2014.
- [20] Kaveh Fathian, Kasra Khosoussi, Parker Lusk, Yulun Tian, and Jonathan P How. CLEAR: A consistent lifting, rembedding, and alignment rectification algorithm for multi-agent data association. *arXiv*, 2019.
- [21] Kristoffer M Frey, Ted J Steiner, and Jonathan P How. Efficient constellation-based map-merging for semantic SLAM. In *IEEE International Conference on Robotics and Automation*, 2019.
- [22] Peng Gao, Brian Reily, Savannah Paul, and Hao Zhang. Visual reference of ambiguous objects for augmented reality-powered human-robot communication in a shared workspace. *International Conference on Virtual, Augmented and Mixed Reality*, 2020.
- [23] Peng Gao, Ziling Zhang, Rui Guo, Hongsheng Lu, and Hao Zhang. Correspondence identification in collaborative robot perception through maximin hypergraph matching. In *International Conference on Robotics and Automation*, 2020.
- [24] Zan Gojcic, Caifa Zhou, Jan D Wegner, and Andreas Wieser. The perfect match: 3D point cloud matching with smoothed densities. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [25] Nitesh B Gundavarapu, Divyansh Srivastava, Rahul Mitra, Abhishek Sharma, and Arjun Jain. Structured Aleatoric Uncertainty in Human Pose Estimation. In *IEEE Conference on Computer Vision and Pattern Recognition, Workshops*, 2019.
- [26] Fei Han, Hua Wang, and Hao Zhang. Learning integrated holism-landmark representations for long-term loop closure detection. In *AAAI Conference on Artificial Intelligence*, 2018.
- [27] Geoffrey A Hollinger, Brendan Englot, Franz Hover, Urbashi Mitra, and Gaurav S Sukhatme. Uncertainty-driven view planning for underwater inspection. In *IEEE International Conference on Robotics and Automation*, pages 4884–4891, 2012.
- [28] Hyunki Hong, Hyeonwoo Yu, and Beom-Hee Lee. Regeneration of normal distributions transform for target lattice based on fusion of truncated Gaussian components. *IEEE Robotics and Automation Letters*, 4(2):684–691, 2019.
- [29] Nan Hu, Qixing Huang, Boris Thibert, and Leonidas J Guibas. Distributable consistent multi-object matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [30] Fabjan Kallasi, Dario Lodi Rizzini, and Stefano Caselli. Fast keypoint features from laser scanner for robot localization and mapping. *IEEE Robotics and Automation Letters*, 1(1):176–183, 2016.
- [31] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems*, 2017.
- [32] Alex Kendall, Vijay Badrinarayanan, and Roberto Cipolla. Bayesian Segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *arXiv*, 2015.
- [33] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [34] Florian Kraus and Klaus Dietmayer. Uncertainty estimation in one-stage object detection. In *IEEE Intelligent Transportation Systems Conference*, pages 53–60, 2019.
- [35] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):453–465, 2013.
- [36] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):453–465, 2014.
- [37] Machine Learning, Stephan Rabanser, Oleksandr Shchur, and Stephan Günnemann. Introduction to Tensor Decompositions and their Applications in Machine Learning. *Machine Learning*, 98(1-2):1–5, 2015.
- [38] Jungmin Lee, Minsu Cho, and Kyoung Mu Lee. Hyper-graph matching via reweighted random walks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

- [39] Spyridon Leonardos, Xiaowei Zhou, and Kostas Daniilidis. Distributed consistent data association via permutation synchronization. In *IEEE International Conference on Robotics and Automation*, 2017.
- [40] Marius Leordeanu and Martial Hebert. A spectral technique for correspondence problems using pairwise constraints. In *IEEE International Conference on Computer Vision*, 2005.
- [41] Qianshan Li, Rong Xiong, and Teresa Vidal-Calleja. A GMM based uncertainty model for point clouds registration. *Robotics and Autonomous Systems*, 91:349–362, 2017.
- [42] Yingzhen Li and Yarin Gal. Dropout inference in Bayesian neural networks with alpha-divergences. In *International Conference on Machine Learning*, 2017.
- [43] Eleonora Maset, Federica Arrigoni, and Andrea Fusiello. Practical and efficient multi-view matching. In *IEEE International Conference on Computer Vision*, 2017.
- [44] James Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, 1957.
- [45] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5):1147–1163, 2015.
- [46] Willie Neiswanger, Chong Wang, and Eric Xing. Asymptotically exact, embarrassingly parallel MCMC. *arXiv preprint arXiv:1311.4780*, 2013.
- [47] H Neudecker. A note on Kronecker matrix products and matrix equation systems. *SIAM Journal on Applied Mathematics*, 17(3):603–606, 1969.
- [48] Andy Nguyen, Mirela Ben-Chen, Katarzyna Welnicka, Yinyu Ye, and Leonidas Guibas. An optimization approach to improving collections of shape maps. In *Computer Graphics Forum*, volume 30, pages 1481–1491, 2011.
- [49] Quynh Nguyen, Antoine Gautier, and Matthias Hein. A flexible tensor block coordinate ascent scheme for hypergraph matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [50] Deepti Pachauri, Risi Kondor, and Vikas Singh. Solving the multi-way matching problem by permutation synchronization. In *Advances in Neural Information Processing Systems*, 2013.
- [51] Brian Reily, Christopher Reardon, and Hao Zhang. Representing Multi-Robot Structure through Multimodal Graph Embedding for the Selection of Robot Teams. *arXiv preprint arXiv:2003.12164*, 2020.
- [52] Cyril Robin and Simon Lacroix. Multi-robot target detection and tracking: taxonomy and survey. *Autonomous Robots*, 40(4):729–760, 2016.
- [53] Madhubhashi Senanayake, Ilankaikone Senthoooran, Jan Carlo Barca, Hoam Chung, Joarder Kamruzzaman, and Manzur Murshed. Search and tracking algorithms for swarms of robots: A survey. *Robotics and Autonomous Systems*, 75:422–434, 2016.
- [54] Héber Sobreira, Carlos M Costa, Ivo Sousa, Luis Rocha, José Lima, PCMA Farias, Paulo Costa, and A Paulo Moreira. Map-matching algorithms for robot self-localization: a comparison between perfect match, iterative closest point and normal distributions transform. *Journal of Intelligent & Robotic Systems*, 93(3-4):533–546, 2019.
- [55] Yumin Suh, Kamil Adamczewski, and Kyoung Mu Lee. Sub-graph matching using compactness prior for robust feature correspondence. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [56] Sebastian Thrun. Probabilistic robotics. *Communications of the ACM*, 45(3):52–57, 2002.
- [57] Yulun Tian, Katherine Liu, Kyel Ok, Loc Tran, Danette Allen, Nicholas Roy, and Jonathan P How. Search and rescue under the forest canopy using multiple UAVs. *arXiv*, 2019.
- [58] Roberto Tron, Xiaowei Zhou, Carlos Esteves, and Kostas Daniilidis. Fast multi-image matching via density-based clustering. In *IEEE International Conference on Computer Vision*, 2017.
- [59] Runzhong Wang, Junchi Yan, and Xiaokang Yang. Learning combinatorial embedding networks for deep graph matching. In *IEEE International Conference on Computer Vision*, 2019.
- [60] ShangGuan Wei, Du Yu, Chai Lin Guo, Liu Dan, and Wang Wei Shu. Survey of connected automated vehicle perception mode: from autonomy to interaction. *IET Intelligent Transport Systems*, 13(3):495–505, 2018.
- [61] Junchi Yan, Zhe Ren, Hongyuan Zha, and Stephen Chu. A constrained clustering based approach for matching a collection of feature sets. In *International Conference on Pattern Recognition*, 2016.
- [62] Zhi Yan, Nicolas Jouandeau, and Arab Ali Cherif. A survey and analysis of multi-robot coordination. *International Journal of Advanced Robotic System*, 10(12):399, 2013.
- [63] Ruonan Zhang and Wenmin Wang. An MCMC-based prior sub-hypergraph matching in presence of outliers. In *International Conference on Pattern Recognition*, 2016.
- [64] Rui Zhao, Wanli Oyang, and Xiaogang Wang. Person re-identification by saliency learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2):356–370, 2016.
- [65] Yiru Zhao, Xu Shen, Zhongming Jin, Hongtao Lu, and Xiansheng Hua. Attribute-driven feature disentangling and temporal aggregation for video person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [66] Xiaowei Zhou, Menglong Zhu, and Kostas Daniilidis. Multi-image matching via fast alternating minimization. In *IEEE International Conference on Computer Vision*, 2015.