Noname manuscript No.
(will be inserted by the editor)

Correspondence Identification for Collaborative Multi-robot Perception under Uncertainty

Peng Gao[1], Rui Guo[2], Hongsheng Lu[2] and Hao Zhang[1]

Received: date / Accepted: date

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 challenging due to the existence of non-covisible objects that cannot be observed by all robots, and due to uncertainty in robot perception. In this paper, we introduce a novel principled approach that formulates correspondence identification as a graph matching problem under the mathematical framework of regularized constrained optimization. We develop a regularization term to explicitly address perception uncertainties by penalizing the object correspondences with a high uncertainty. We also introduce a second regularization term to explicitly address non-covisible objects by penalizing the correspondences built by the non-covisible objects. Our approach is evaluated in robotic simulations and real physical robots. Experimental results show that our method is able to address correspondence identification under uncertainty and non-covisibility, and achieves the state-of-the-art performance.

Keywords Correspondence identification · Regularized graph matching · Perceptual uncertainty · Collaborative perception

Peng Gao · Hao Zhang
Colorado School of Mines
E-mail: {gaopeng, hzhang}@mines.com
Address: 1500 Illinois St, Golden, CO 80401

Rui Guo · Hongsheng Lu
Toyota Motor North America, InfoTech Laboratory.
E-mail: { rui.guo, hongsheng.lu}@toyota.com
Address: 465 N Bernardo Ave, Mountain View, CA 94043

1 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 (Brambilla, Ferrante, Birattari, and Dorigo, 2013; Chung, Paranjape, Dames, Shen, and Kumar, 2018; Yan, Jouandeau, and Cherif, 2013). 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 (Lampert, Nickisch, and Harmeling, 2014; Robin and Lacroix, 2016; Senanayake, Senthooran, Barca, Chung, Kamruzzaman, and Murshed, 2016; Reily, Reardon, and Zhang, 2020), connected and autonomous driving (Wei, Yu, Guo, Dan, and Shu, 2018), collaborative manufacture (Dogar, Spielberg, Baker, and Rus, 2019), and multi-robot localization and mapping (Aragues, Montijano, and Sagues, 2011; Nguyen, Ben-Chen, Welnicka, Ye, and Guibas, 2011).

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 (Frey, Steiner, and How, 2019; Kallasi, Rizzini, and Caselli, 2016; Leonardos, Zhou, and Daniilidis, 2017; Tian, Liu, Ok, Tran, Allen, Roy, and How, 2019; Gao, Reily, Paul, and Zhang, 2020b). 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 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

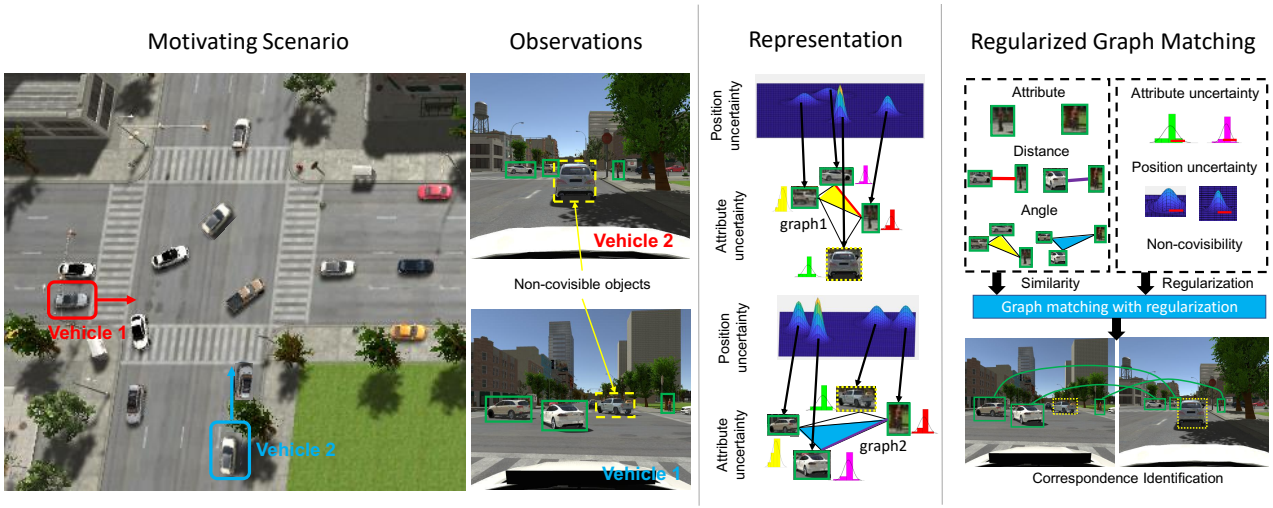


Fig. 1 A motivating example of correspondence identification with non-covisible objects under uncertainty for multi-robot collaborative perception. Before two connected autonomous vehicles share information of street objects, they must identify the correspondence of these objects, while addressing non-covisible objects and perception uncertainty, in order to correctly refer to the same objects from their own fields of view. Given a pair of observations with detected objects, our approach performs regularized graph matching for correspondence identification of objects. Specifically, we represent each observation as a graph. Each node denotes an object's attribute and edges describe the spatial relationships among objects. Each correspondence is identified based on the attribute, and spatial (including distance and angle) similarities. Meanwhile, we design a new regularization term to penalize the correspondence constructed by the non-covisible objects, and 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.

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 (Chen, Zhu, and Gong, 2017; Gojcic, Zhou, Wegner, and Wieser, 2019). The second category of methods employ synchronization algorithms to identify circle-consistent associations among objects (Aragues, Montijano, and Sagues, 2011; Fathian, Khosoussi, Lusk, Tian, and How, 2019; Hu, Huang, Thibert, and Guibas, 2018; Tron, Zhou, Esteves, and Daniilidis, 2017), which synchronize associations of the same objects observed from multiple views. The third category of methods are based on spatial (e.g., distance or angular) and geometric information, e.g., using linear assignment (Munkres, 1957), quadratic assignment (Cho, Lee, and Lee, 2010; Leordeanu and Hebert, 2005), and high-order graph matching (Chang, Fischer, Petit, Zambelli, and Demiris, 2017; Duchenne, Bach, Kweon, and Ponce, 2011; Nguyen, Gautier, and Hein, 2015).

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 (Kendall, Gal, and Cipolla, 2018)) 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 uncer-

tainties 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.

A preliminary conference version of this work was published at Robotics Science and System 2020 (Gao, Guo, Lu, and Zhang, 2020a). We extend the previous conference work as follows. In Section 3.1.1 and Section 3.1.2, we introduce our feature extraction process for correspondence identification. In Section 3.3.1 and Section 3.3.2, we introduce a uncertainty quantification technique based on Monte Carlo dropout to quantify attribute and position uncertainties for the regularized graph matching. In Section 3.5, we design a new approach for multi-robot collaborative object localization based on the identified correspondences. In Section 4.5, we perform a case study by implementing our approach with a physical multi-robot system in the scenario of multi-robot collaborative object localization.

2 Related Work

We review existing techniques for correspondence identification in multi-robot collaborative perception. Specially, we discuss the category of the existing correspondence identification techniques, the definition of the traditional perception uncertainty and the uncertainty quantification techniques in robotic perception.

2.1 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 intro 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 (Engel, Schöps, and Cremers, 2014), ORB (Mur-Artal, Montiel, and Tardos, 2015) and 3D keypoints (Boroson and Ayanian, 2019). Furthermore, to identify same individual objects with changing appearances, visual features (Zhao, Oyang, and Wang, 2016) and attribute features (Zhao, Shen, Jin, Lu, and Hua, 2019) are used for re-identification.

Synchronization-based identification recognizes correspondences of objects from multiple views by satisfying the circle-consistent constraint (Fathian, Khosoussi, Lusk, Tian, and How, 2019). 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 (Boyd, Parikh, Chu, Peleato, Eckstein et al., 2011), which can be relaxed to be convex (Hu, Huang, Thibert, and Guibas, 2018) and solved using a convex optimization solver (Zhou, Zhu, and Daniilidis, 2015). 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 (Maset, Arrigoni, and Fusiello, 2017; Pachauri, Kondor, and Singh, 2013). The graph clustering methods formulate the multi-view object correspondence problem as graph clustering that is solved, e.g., by graph cut (Fathian, Khosoussi, Lusk, Tian, and How, 2019) or k-means (Yan, Ren, Zha, and Chu, 2016).

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 (Sobreira, Costa, Sousa, Rocha, Lima, Farias, Costa, and Moreira, 2019). Correspondence identification is also formulated as a linear assignment problem solved by the Hungarian (Almohamad and Duffuaa, 1993) or Sinkhorn algorithm (Wang, Yan, and Yang, 2019), and a quadratic assignment problem that considers distances between the objects (Cho, Lee, and Lee, 2010; Leordeanu and Hebert, 2005). 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 (Chang, Fischer, Petit, Zambelli, and Demiris, 2017; Lee, Cho, and Lee, 2011), tensor block coordinate ascent (Nguyen, Gautier, and Hein, 2015), and Monte Carlo Markov Chain (MCMC) (Suh, Adamczewski, and Mu Lee, 2015).

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.

2.2 Uncertainty in Perception

Uncertainty in robot perception is traditionally computed as the variance of probability distributions (Thrun, 2002). 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 (Black and Ellis, 2002), modeling odometry uncertainty with fuzzy set for robot motion estimation (Buschka, Saffiotti, and Wasik, 2000), adding uncertainty to robot joint positions to improve reliability estimation (Carreras and Walker, 2001), describing uncertainty in point clouds by Gaussian Mixture Model (GMM) to compensate distortion (Hong, Yu, and Lee, 2019; Li, Xiong, and Vidal-Calleja, 2017), applying a multi-variate Gaussian distribution to model human joints to improve pose prediction (Gundavarapu, Srivastava, Mitra, Sharma, and Jain, 2019), and modeling uncertainty of point cloud positions by regression for robot inspection (Hollinger, Englot, Hover, Mitra, and Sukhatme, 2012). 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 (Kendall, Gal, and Cipolla, 2018), camera localization (Bertoni, Kreiss, and Alahi, 2019; Kendall, Gal, and Cipolla, 2018), and object classification (Kraus and Dietmayer, 2019). Uncertainty in machine perception based on BNN is generally divided in two categories (Der Kiureghian and Ditlevsen, 2009; Kendall and Gal, 2017): 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 (Kendall and Gal, 2017) (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 approximate Bayesian inference.

In order to perform uncertainty quantification under Bayesian framework, traditional methods focus on modeling the posterior distribution over model parameters to obtain the epistemic uncertainty (Neal, 2012; Richard and Lippmann, 1991). Since the posterior distributions are intractable, variational inference (Li and Gal, 2017) is proposed to approximate the posterior distribution of model parameters, such as Markov chain Monte Carlo (MCMC) (Ding, Fang, Babbush, Chen, Skeel, and Neven, 2014), expectation propagation (Li, Hernández-Lobato, and Turner, 2015), stochastic gradient MCMC (Korattikara Balan, Rathod, Murphy, and Welling, 2015) and Monte Carlo dropout (Kendall, Badrinarayanan, and Cipolla, 2015a).

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.

3 The Proposed Approach

We present our novel regularized graph matching approach for correspondence identification in collaborative perception, which explicitly addresses perception uncertainty and non-covisibile objects in a unified mathematical framework. Specifically, we discuss our feature extraction, problem formulation, addressing perception uncertainty and non-covisibility.

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.

3.1 Feature Extraction

Compared with our previous work (Gao, Guo, Lu, and Zhang, 2020a), we discuss the details on extracting appearance and spatial features of objects from monocular observations (images) to perform multi-robot correspondence identification. An object's appearance is represented as a distribution of its classification category, which is obtained from the instance-level segmentation. In addition, the location of an object is obtained from the monocular depth estimation technique.

3.1.1 Instance Segmentation

Instance segmentation is a problem of detecting and locating individual objects of interest appearing in images. Given

an image observation with n objects, we perform the mask-RCNN (He, Gkioxari, Dollár, and Girshick, 2017) to map the raw observation into a 3-channel outputs, including the classes, bounding boxes and masks of objects of interest. Formally, the loss functions are defined as follows:

- The loss function for object classification is based on cross entropy, which is defined as $L_c = -\hat{a_i^c} \log a_i^c$, where $a_i^c \in \mathbb{R}$ denotes the confidence of the i-th object belonging to the c-th class, which is obtained by SoftMax, and $\hat{a_i^c}$ denotes the ground true confidence.
- The loss function for bounding box regression is defined as $L_b = f(\mathbf{b}_i - \hat{\mathbf{b}_i})$, where $\mathbf{b}_i = [x^{2d}, y^{2d}, w^{2d}, h^{2d}]$ denotes the predicted bounding box of the i-th object with central point x^{2d}, y^{2d}, width w^{2d} and height h^{2d} in 2D image coordination, f denotes the smooth function (Girshick, 2015).
- The loss function for mask generation is defined as $L_m = \frac{1}{K}\sum_{k=1}^{K}(\mathbf{m}_i \circ \hat{\mathbf{m}}_i)$, where $\circ$ denotes the element-wise product, K is the number of pixels of the i-th object, $\mathbf{m}_i$ denotes the predicted mask and $\hat{\mathbf{m}}_i$ denotes the ground truth in a binary form.

The final loss function is defined as $L_{is} = L_c + L_b + L_m$. Minimising the loss function is equivalent to minimize the error of the object classification, bounding box regression and mask generation.

3.1.2 Monocular Depth Estimation

Monocular depth estimation is a task of estimating the depth of the scene given a single image. Given an image observation, we perform the CNN-based depth estimation network (Hu, Zhang, and Okatani, 2019) to estimate the pixel-wise depth of individual objects. In the network, ResNet-50 (Laina, Rupprecht, Belagiannis, Tombari, and Navab, 2016) is used to extract high-level features of raw images. The loss function is defined as $L_{de} = ||\hat{z} - z||_2^2$, where z denotes the pixel-level depth value and $\hat{z}$ denotes the ground true depth.

Given the results obtained from depth estimation and the instance segmentation, we represent the attribute feature and the 3D position of objects as follows:

- We represent the attribute feature of an object as $\mathbf{a}_i = [a_i^1, a_i^2, \ldots, a_i^m]$, where $a_i^c, c = 1, 2, \ldots, m$ denotes the confidence of the i-th object belonging to the c-th class and m denotes the number of categories for object classification.
- We represent the 3D position of an object as $\mathbf{p} = [x, y, z]$, where z denotes the averaged depth of the object, which is calculated by averaging all the pixel-wise depth values of the object given its mask. x and y denotes the location of the object in real-world coordination, which is transformed from the 2D central point x^{2d}, y^{2d} based on the intrinsic camera parameters (Szeliski, 2010).

3.2 Problem Formulation

Based upon the attribute features and 3D positions of objects, 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 $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$ 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. $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ 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 $s_{i,j}^d, i, j = 1, 2, \ldots, n, i \neq j$ 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 $s_{i,j,k}^a = [\theta_i, \theta_j, \theta_k], i, j, k = 1, 2, \ldots, n, i \neq j \neq k$ 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 more robust to geometric variations (e.g., the deformation of spatial relationships of objects, which is caused by sensor noise) compared to the distance relationship (Duchenne, Bach, Kweon, and Ponce, 2011).

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.

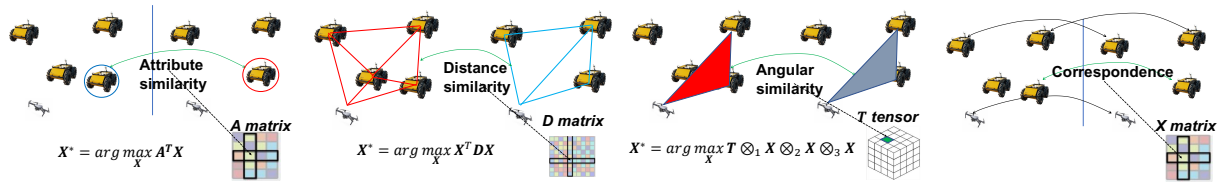
- The attribute similarity is computed by

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

  where $\mathbf{A}_{ii'}$ 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

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

  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 γ 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).



(a) Attribute similarity matrix (b) Distance similarity matrix (c) Angular similarity tensor (d) Correspondence matrix

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 (Gao, Guo, Lu, and Zhang, 2020a).

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

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

where $\mathbf{T}_{ii',jj',kk'}$ denotes the similarity between triangle $s^a_{i,j,k} \in \mathcal{S}^a$ and triangle $s^{a'}_{i',j',k'} \in \mathcal{S}^{a'}$. 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:

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

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 $\mathbf{T}$ (Learning, Rabanser, Shchur, and Günnemann, 2015), 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.

3.3 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. Compared with our previous work (Gao, Guo, Lu, and Zhang, 2020a), we explicitly quantify attribute and position uncertainties by designing segmentation and depth estimation neural networks under Bayesian framework in this paper.

3.3.1 Attribute uncertainty representation

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. We can utilize Bayesian neural networks (BNN) (Kendall and Gal, 2017) to directly estimate the attribute uncertainty. Formally, we formulate the instance segmentation neural network (mask-RCNN) from the Bayesian perspective as follows:

$$p(a_i^c|\mathbf{I}_i,\mathcal{T}) = \int_{\mathbf{W}^s\in\Omega} p(a_i^c|\mathbf{I}_i,\mathbf{W}^s)p(\mathbf{W}^s|\mathcal{T})d\mathbf{W}^s \tag{5}$$

where $\mathbf{I}_i$ denotes the i-th image region that contains an object, a_i^c denotes the confidence the i-th image region belonging to the c-th category, $\mathcal{T}$ denotes the training dataset with input images and corresponding ground truth classes, and $\mathbf{W}^s$ denotes the trainable parameter of the mask-RCNN in a distribution form instead of taking fixed values. The parameter matrix $\mathbf{W}^s$ can be obtained by calculating the posterior distribution of Eq. (5)

Since Eq. (5) is intractable (Li and Gal, 2017), we apply Monte Carlo dropout sampling to obtain the distribution-form parameter matrix $\mathbf{W}^s$ by approximating the posterior distribution. During training, by adding dropout technique to the parameter matrix $\mathbf{W}^s$, several parameters are dropped

out based on Bernoulli distribution. When performing inference, we also enable dropout to sample $\mathbf{W}^s$. Finally, the distribution-form attribute inference is defined as follows:

$$p(a_i^c|\mathbf{I}_i,\mathcal{T}) \approx \frac{1}{T}\sum_{t=1}^{T} p(a_i^{c\{t\}}|\mathbf{I}_i,\mathbf{W}^s), \mathbf{W}^{(s)} \sim q(\mathbf{W}^s) \quad (6)$$

where T is the number of samplings, $q(\mathbf{W}^s)$ is the approximated posterior distribution, which is obtained by minimizing the Kullback-Leibler divergency (Kendall and Gal, 2017).

We define the final classification result as the expectation of the samplings sampled from Eq. (6). 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 aleotoric 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 $\mathbf{v} = [v_1, v_2, \ldots, v_n]$ 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:

$$\mathbf{V} = \mathbf{v} \oplus \mathbf{v}'^{\top} \quad (7)$$

where $\oplus$ denotes the kronecker plus (Neudecker, 1969), 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 $\mathbf{V}$:

$$\mathbf{W}^a = \exp\left(-\frac{1}{\sigma}\mathbf{V}\right) \quad (8)$$

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

3.3.2 Position uncertainty representation

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, we apply Monte Carlo dropout technique to our depth estimation network, the provided uncertainty is a combination of the epistemic uncertainty in the BNN model and the aleotoric uncertainty in the data.

Given the position uncertainties of objects, we define $\mathbf{u} = [u_1, u_2, \ldots, u_n]$ to represent the position uncertainties of all objects encoded by graph $\mathcal{G}$. Given $\mathbf{u}$ and $\mathbf{u}'$ from $\mathcal{G}$ and $\mathcal{G}'$ respectively, the position uncertainty matrix $\mathbf{U}$ is computed by:

$$\mathbf{U} = \mathbf{u} \oplus \mathbf{u}'^{\top} \quad (9)$$

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 $s^d_{i,j} \in \mathcal{S}^d$ and distance $s^{d'}_{i',j'} \in \mathcal{S}^{d'}$ 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:

$$\mathbf{W}^d = \exp\left(-\frac{1}{\sigma}\left(\mathbf{U} \oplus \mathbf{U}^{\top}\right)\right) \quad (10)$$

where $\mathbf{W}^d = \{\mathbf{W}^d_{ii',jj'}\} \in \mathbb{R}^{nn'\times nn'}$, and $\mathbf{W}^d_{ii',jj'}$ 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:

$$\mathbf{W}^t = \exp\left(-\frac{1}{\sigma}\left(\mathbf{U} \oplus \left(\mathbf{U} \oplus \mathbf{U}^{\top}\right)^{\top}\right)\right) \quad (11)$$

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

3.3.3 Addressing uncertainty

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:

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

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:

$$\begin{aligned}\max_{\mathbf{X}}\ & (\mathbf{A}\circ\mathbf{W}^a)^\top\mathbf{x}+\mathbf{x}^\top(\mathbf{D}\circ\mathbf{W}^d)\mathbf{x}\\ &+\mathbf{T}\circ\mathbf{W}^t\otimes_1\mathbf{x}\otimes_2\mathbf{x}\otimes_3\mathbf{x}-\lambda_1||(\mathbf{V}+\mathbf{U})\circ\mathbf{X}||^2\\ \text{s.t.}\quad & \mathbf{X}\mathbf{1}_{n'\times1}\le\mathbf{1}_{n\times1},\mathbf{X}^\top\mathbf{1}_{n\times1}\le\mathbf{1}_{n'\times1}\end{aligned}\tag{13}$$

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 λ_1 is introduced to balance the maximization of the overall similarity and the minimization of the overall uncertainty.

3.3.4 Addressing 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:

$$\begin{aligned}\max_{\mathbf{X}}\ & (\mathbf{A}\circ\mathbf{W}^a)^\top\mathbf{x}+\mathbf{x}^\top(\mathbf{D}\circ\mathbf{W}^d)\mathbf{x}\\ &+\mathbf{T}\circ\mathbf{W}^t\otimes_1\mathbf{x}\otimes_2\mathbf{x}\otimes_3\mathbf{x}-\lambda_2||\mathbf{X}||^2\\ \text{s.t.}\quad & \mathbf{X}\mathbf{1}_{n'\times1}\le\mathbf{1}_{n\times1},\mathbf{X}^\top\mathbf{1}_{n\times1}\le\mathbf{1}_{n'\times1}\end{aligned}\tag{14}$$

where λ_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:

$$\begin{aligned}\max_{\mathbf{X}}\ & (\mathbf{A}\circ\mathbf{W}^a)^\top\mathbf{x}+\mathbf{x}^\top(\mathbf{D}\circ\mathbf{W}^d)\mathbf{x}\\ &+\mathbf{T}\circ\mathbf{W}^t\otimes_1\mathbf{x}\otimes_2\mathbf{x}\otimes_3\mathbf{x}\\ &-\lambda_1||(\mathbf{V}+\mathbf{U})\circ\mathbf{X}||^2-\lambda_2||\mathbf{X}||^2\\ \text{s.t.}\quad & \mathbf{X}\mathbf{1}_{n'\times1}\le\mathbf{1}_{n\times1},\mathbf{X}^\top\mathbf{1}_{n\times1}\le\mathbf{1}_{n'\times1}\end{aligned}\tag{15}$$

3.4 Optimization Algorithm

Since the proposed constrained optimization formulation in Eq. (15) is a non-convex problem and has regularization terms, the commonly used optimization methods for graph matching, e.g., based upon reweighted random walks (Chang, Fischer, Petit, Zambelli, and Demiris, 2017; Lee, Cho, and Lee, 2011), cannot be directly utilized to solve the problem. Thus, we design a new heuristic optimization algorithm based on Markov chain Monte Carlo (MCMC) sampling (Suh, Adamczewski, and Mu Lee, 2015).

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. (15) to the following:

$$\begin{aligned}P(\mathbf{y})=\exp((&\mathbf{A}\circ\mathbf{W}^a)^\top\pi(\mathbf{y})+\pi(\mathbf{y})^\top(\mathbf{D}\circ\mathbf{W}^d)\pi(\mathbf{y})\\ &+\mathbf{T}\circ\mathbf{W}^t\otimes_1\pi(\mathbf{y})\otimes_2\pi(\mathbf{y})\otimes_3\pi(\mathbf{y})\\ &-\lambda_1||(\mathbf{U}+\mathbf{V})\circ\pi(\mathbf{y})||^2)-\lambda_2||\pi(\mathbf{y})||^2\end{aligned}\tag{16}$$

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:

$$\begin{aligned}\pi(\mathbf{y})=\max_{\mathbf{X}}\quad & (\mathbf{A}\circ\mathbf{W}^a)^\top\mathbf{x}+\mathbf{x}^\top(\mathbf{D}\circ\mathbf{W}^d)\mathbf{x}\\ &+\mathbf{T}\circ\mathbf{W}^t\otimes_1\mathbf{x}\otimes_2\mathbf{x}\otimes_3\mathbf{x}\\ \text{s.t.}\quad & \sum_i\mathbf{X}_{ij}=1,\text{if }\mathbf{y}_i=1\\ & \sum_j\mathbf{X}_{ij}=0,\text{if }\mathbf{y}_i=0\\ & \mathbf{X}\mathbf{1}_{n'\times1}\le\mathbf{1}_{n\times1},\mathbf{X}^\top\mathbf{1}_{n\times1}\le\mathbf{1}_{n'\times1}\end{aligned}\tag{17}$$

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 $\mathbf{y}$:

$$q(\mathbf{y},\hat{\mathbf{y}})=\begin{cases}\alpha\ \exp\left(-\frac{1}{\gamma}\left(\hat{\mathbf{y}}^\top(\mathbf{v}+\mathbf{u})\right)\right) & ||\hat{\mathbf{y}}||_1=||\mathbf{y}||_1+1\\ (1-\alpha)\frac{1}{||\mathbf{y}||_1} & ||\hat{\mathbf{y}}||_1=||\mathbf{y}||_1-1\\ 0 & otherwise\end{cases}\tag{18}$$

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

Algorithm 1: The proposed solver to solve the formulated non-convex regularized constrained optimization problem in Eq. (15).

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: $\mathbf{x} \in \{0,1\}^{nn'}$

Initialize $\mathbf{y} \in \{0,1\}^n$ (belongs to $\mathcal{G}$) and T;
while $T > T_f$ **do**
  Compute the active node distribution $q(\mathbf{y} \to \hat{\mathbf{y}})$ in Eq. (18);
  Sample $\mathbf{y}$ from q;
  Calculate the acceptance ratio $\alpha(\mathbf{y} \to \hat{\mathbf{y}})$ in Eq. (19);
  if $\alpha(\mathbf{y} \to \hat{\mathbf{y}}) > \epsilon$ **then**
    $\mathbf{y} \to \hat{\mathbf{y}}$;
    if $P(\hat{\mathbf{y}}) > P(\mathbf{y})$ *in Eq. (16)* **then**
      Calculate $\mathbf{x}^*$ in Eq. (17);
      Update state $\mathbf{y} = \hat{\mathbf{y}}$;
    end
  end
  $T = \xi T$;
end
return $\mathbf{x}^*$

deleting one active node $||\hat{\mathbf{y}}||_1 = ||\mathbf{y}||_1 - 1$; otherwise the probability equals to 0. The mode selection is controlled by α, 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 (Brooks, Gelman, Jones, and Meng, 2011), which means the designed Markov chain is reversible. According to (Suh, Adamczewski, and Mu Lee, 2015), the acceptance ratio of state update is designed as:

$$\alpha(\mathbf{y}, \hat{\mathbf{y}}) = \min\left(\frac{P(\hat{\mathbf{y}})q(\hat{\mathbf{y}}, \mathbf{y})}{P(\mathbf{y})q(\mathbf{y}, \hat{\mathbf{y}})}, 1\right) \tag{19}$$

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 ξ.

Complexity. The complexity to solve the optimization problem in Eq. (13) 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 $O(n^4)$.

3.5 Multi-robot Collaborative Object Localization based on Identified Correspondences

Based on the identified correspondences of objects, we further integrate multi-robot observations of the same object to improve the accuracy of the measured positions of objects (a.k.a, object localization). Specifically, given a multi-robot system with N robots ($N > 2$), each robot has its own observations with detected objects. By performing the algorithm 1 among pairs of observations, we identify the correspondences between pairs of observations. Given the identified pairwise correspondences, post-processing can be applied to aggregate pairwise correspondence results by forcing circle consistency Hu et al. (2018).

We assume that M objects are identified based on Algorithm 1 and these objects can be observed by all the N robots. We define the measured positions of objects as $\mathcal{P}^n = \{\mathbf{p}_1^n, \mathbf{p}_2^n, \ldots, \mathbf{p}_M^n\}, n = 1, 2, \ldots, N$, where $\mathbf{p}_i^n$ denotes the measured position of the i-th object obtained from the n-th robot. Based on the identified correspondences, $\mathbf{p}_k^i \in \mathcal{P}^i$ and $\mathbf{p}_k^j \in \mathcal{P}^j$ denote the measured positions of the same object obtained by both i-th and j-th robots. The position uncertainties of objects are represented as $\mathcal{U}^n = \{u_1^n, u_2^n, \ldots, u_M^n\}$, where u_i^n denotes the position uncertainty of $\mathbf{p}_i^n$.

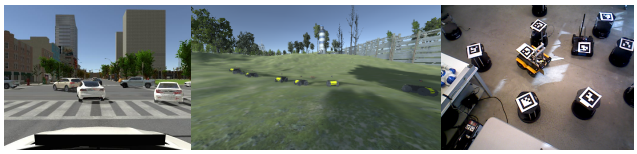
In order to improve the accuracy of single-robot measured positions of objects, we propose a multi-robot fusion gain to integrate multi-robot measurements, which is defined as follows:

$$\mathbf{M}_i^n = \left(\sum_{j=1}^{N} (\mathbf{I}u_i^j)^{-1}\right)^{-1} (\mathbf{I}u_i^n)^{-1} \tag{20}$$

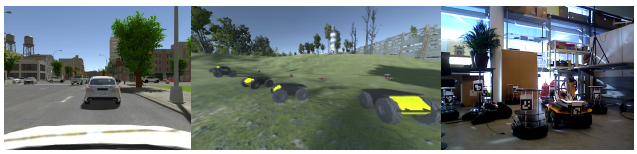
where $\mathbf{M}_i^n \in \mathcal{R}^{3\times3}$ denotes the measurement fusion gain for the n-th robot's measurement of the i-th object and $\mathbf{I} \in \mathcal{R}^{3\times3}$ denotes an identity matrix. In addition, $\mathbf{M}_i^n$ follows the constraint $\sum_{n=1}^{N} \mathbf{M}_i^n = \mathbf{I}$. The fusion gain for each robot represents the weight of each robot's measurement in all the multi-robot measurements given the normalized position uncertainties. The final position estimation of an object is defined as follows:

$$\hat{\mathbf{p}}_i^n = \mathbf{M}_i^n \mathbf{p}_i^n + \sum_{j=1, j\neq n}^{N} \mathbf{M}_i^j \sigma(\mathbf{p}_i^j) \tag{21}$$

where σ denotes the transformation function that transforms the multi-robot measured positions to the n-th robot's coordinates based on camera extrinsic parameters (Zhang and Pless, 2004). The camera extrinsic parameters can be obtained through GPS or deep learning algorithm (Kendall,



(a) Observations by one robot



(b) Observations by the other robot

Fig. 3 Illustrations of CAD, S-MRC, and R-MRC scenarios (Gao, Guo, Lu, and Zhang, 2020a).

Grimes, and Cipolla, 2015b). $\hat{\mathbf{p}}_i^n$ denotes the final position estimation of the i-th object observed by the n-th robot, which is computed by the sum of single-robot measured positions weighted by the fusion gains $\mathbf{M}_i^n$. If a robot's measured position of an object has large uncertainty (e.g., perception uncertainty caused by occlusion), then its contribution will be heavily weakened during the fusion. The uncertainty of the final position estimation is defined as follows:

$$\hat{u}_i^n = \left(\sum_{n=1}^{N} (u_i^n)^{-1} \right)^{-1} \tag{22}$$

where $\hat{u}_i^n$ denotes the uncertainty of the final position estimation $\hat{\mathbf{x}}_i^n$, which is obtained by integrating all the single-robot position uncertainties.

4 Experiment

4.1 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)[1]. 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). The QR code are used only as ground truth for evaluation, but not used as input in the experiment.

[1] The datasets are available at: `http://hcr.mines.edu/project/civr.html`.

We use mask-RCNN to do the instance segmentation from the raw images, the results are presented in Figure 4(b). We use the attribute feature constructed from the distribution of classification results, which is defined in Section 3.1.1. Each element in the attribute feature vector denotes an confidence of the object belonging to a specific class. The uncertainty of the attribute feature are obtained from BNN models, which is defined in Section 3.3.1. The attribute uncertainties are presented in Figure 5(a). Object positions are computed from depth estimation, which is defined in Section 3.1.2, the results are shown in Figure 4(c). The position uncertainties are defined in Section 3.3.2, which are shown in Figure 5(b).

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 λ_2 is greater than λ_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.

We adopt precision and recall as metrics to evaluate the performance of correspondence identification, following (Suh, Adamczewski, and Mu Lee, 2015; Zhang and Wang, 2016). 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 covisible objects over all ground truth correspondences of covisible objects.

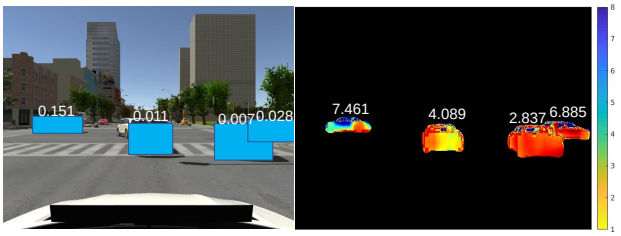
Also, for comparison, we implement six previous correspondence identification techniques based on object appearance or spatial information. In terms of the spatial-based methods, we implement (1) pairwise graph matching **RRWM** (Cho, Lee, and Lee, 2010) that uses the distance similarity to identify correspondences, (2) iterative closest point **ICP** (Besl and McKay, 1992) that iteratively minimizes the distances of two graphs, two hypergraph matching techniques, including (3) **BCAGM** (Nguyen, Gautier, and Hein, 2015) and (4) **RRWHM** (Lee, Cho, and Lee, 2011) 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**) (Zhao, Shen, Jin, Lu, and Hua, 2019), and (6) an approach based upon multi-order similarities (**MOS**) (Chang, Fischer,

(a) Image observations

(b) Instance segmentation

(c) Depth estimation

Fig. 4 Illustration of the image observed by a vehicle in CAD (first row), instance segmentation (second row) and depth estimation (third row). In Figure 4(c), a darker color within an object region indicates a closer distance.



(a) Attribute uncertainty (b) Position uncertainty

Fig. 5 Illustration of the attribute uncertainty and position uncertainty. The numbers in Figures 5(a) and 5(b) denote average attribute and position uncertainties for each object, respectively. (Gao, Guo, Lu, and Zhang, 2020a).

Petit, Zambelli, and Demiris, 2017) that consider both appearance and spatial relationships.

4.2 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 1. 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 1. 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 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 6(c), which demonstrates that our approach correctly identifies correspondences of street objects. For

Table 1 Quantitative results of our approach, and comparisons with previous and baseline methods in CAD, S-MRC and R-MRC. The results are presented as the mean value, which are computed by running these methods five times in each scenario. (Gao, Guo, Lu, and Zhang, 2020a).

Method	CAD		S-MRC		R-MRC	
	Precision	Recall	Precision	Recall	Precision	Recall
RRWM	0.032	0.029	0.289	0.291	0.131	0.106
ICP	0.061	0.069	0.302	0.311	0.089	0.116
BCAGM	0.154	0.186	0.503	0.587	0.201	0.256
RRWHM	0.161	0.173	0.524	0.592	0.219	0.221
ReId	0.238	0.264	0.321	0.472	0.396	0.508
MOS	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
Ours	**0.659**	**0.718**	**0.701**	**0.711**	**0.575**	**0.723**

comparison, we also include the qualitative results obtained by ReId and MOS in Figure 6(a) and Figure 6(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 7. 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 (Neiswanger, Wang, and Xing, 2013) is applied, the execution speed can be further improved to around 40Hz on an 8-cores CPU.

4.3 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.

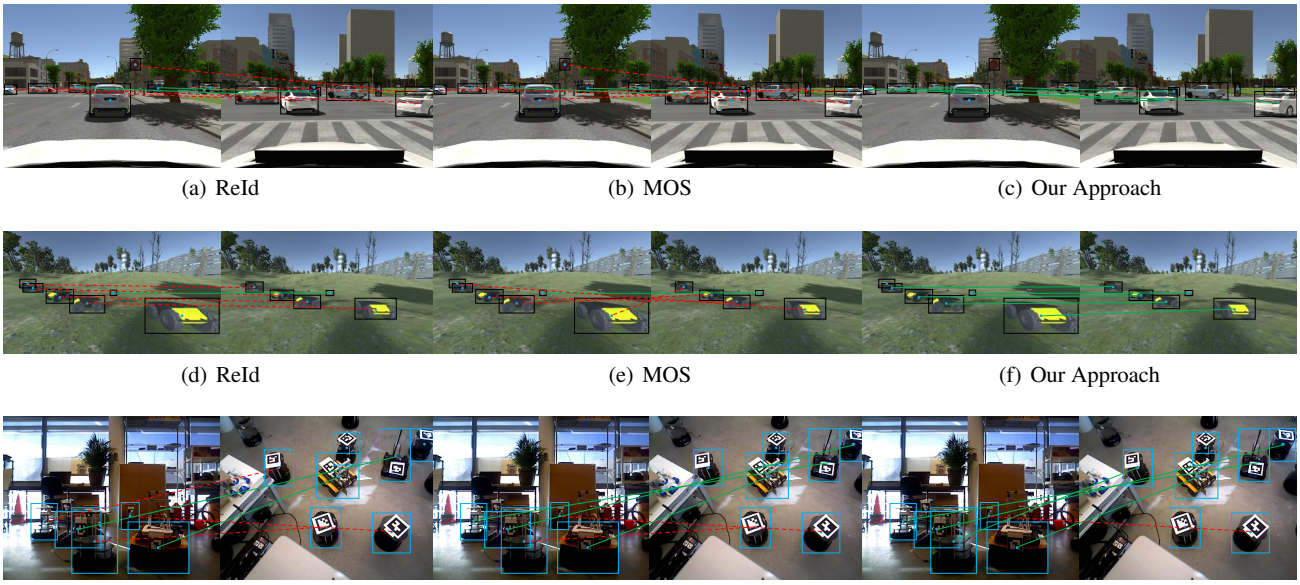
The qualitative results in S-MRC are shown in Figure 6(f). It demonstrates that our approach is able to correctly identifies correspondences of the UGVs from two views. Comparisons with ReId and MOS are also presented in Figure 6. We observe that ReId does not work well since most objects are identical; 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 1. 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 1. 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.

4.4 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 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 6. 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

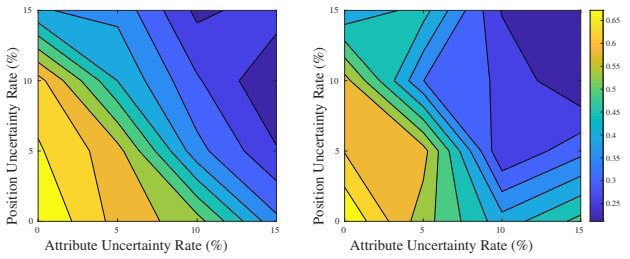


(a) ReId (b) MOS (c) Our Approach

(d) ReId (e) MOS (f) Our Approach

(g) ReId (h) MOS (i) Our Approach

Fig. 6 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 solid lines denote correct correspondences, red dash lines denote incorrect correspondences (Gao, Guo, Lu, and Zhang, 2020a).



(a) Precision (b) Recall

Fig. 7 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 (Gao, Guo, Lu, and Zhang, 2020a).

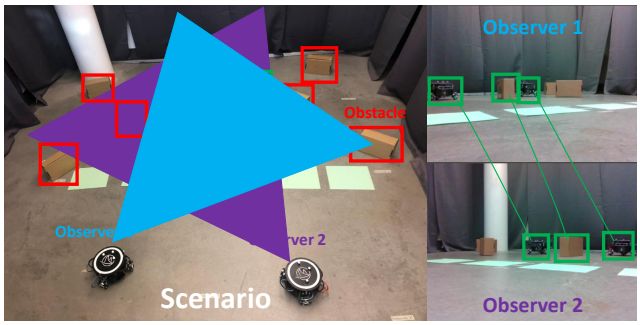


Fig. 8 Illustrations of correspondence identification in the scenario of multi-robot collaborative object localization. When two collaborative robots (observer 1 and 2) collaboratively localize objects (target robots or obstacles), they need to identify the correspondences of objects to refer to the same object.

of correspondence identification in R-MRC are listed in Table 1. 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.

4.5 Case Study with Physical Multi-robot System

We finally perform a case study of our approach by implementing it with a physical multi-robot system. The system consists of four mobile robots (two observers and two dynamic targets) equipped with the RealSense D435 camera. Each mobile robot is powered by a dual core 2.9GHz CPU and lacks an onboard GPU. The ground truth locations and correspondences are obtained by an Optitrack motion tracking system.

Similar to the settings in Section 4.3, the mobile robots and obstacles used in this scenario have identical appearances, as shown in Figure(8). The number of non-covisible objects is twice as the number of objects that can be ob-

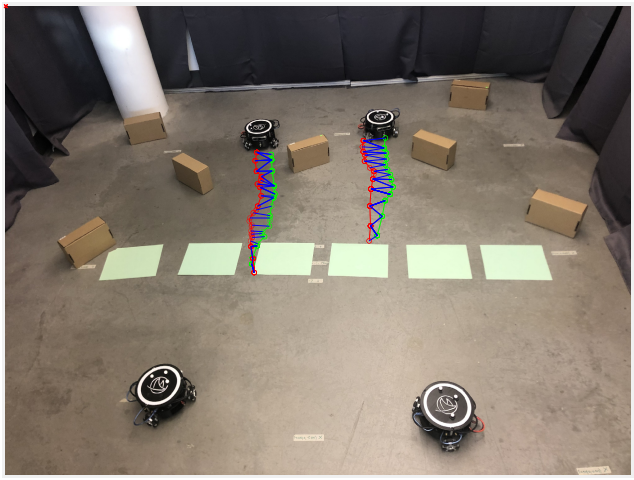


Fig. 9 Qualitative results of our approach in the scenario of multi-robot collaborative object localization. The red curve denotes the trajectory recorded by observer 1, the green curve denotes the trajectory recorded by observer 2, and the blue line denotes the correct correspondences of the target robots identified between both observations.

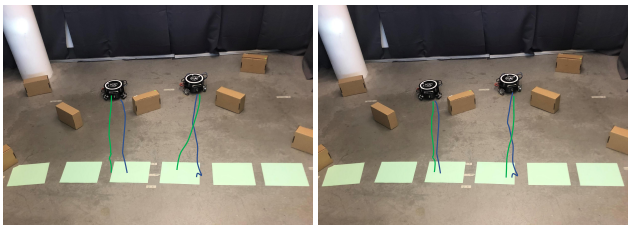
served by both robots. In addition, there exist strong occlusion in robot's observations, which leads to large perception uncertainty.

The qualitative results of our approach are presented in Figure (9). Given the results, we can see that our approach can well identify the correspondences of target robots between two robots' fields of views. In addition, our approach achieves 0.9986 precision and 0.7268 recall in this challenging scenario with large number of non-covisible objects and strong perception uncertainty..

Based on the identified correspondences, we further evaluate our proposed approach of multi-robot collaborative object localization. Specifically, we compared our proposed multi-robot collaborative object localization approach with the baseline method that only uses single-robot observations. The qualitative results are shown in Figure 10. Compared with single-robot observations, our approach significantly improve the object localization results by integrating multi-robot observations. We use displacement error to be the evaluation metrics, which is defined as the Euclidean distance between the estimated locations and the ground truth locations. Based on the metrics, the baseline method obtains 17.5322mm displacement error and our approach obtains 12.0441mm displacement error, which significantly improves the performance of object localization.

5 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



(a) Single-robot Observations (b) Our Approach

Fig. 10 Qualitative results of our approach in the scenario of multi-robot collaborative object localization. The blue curve denotes the ground truth trajectory and the green curve denotes the estimated trajectory.

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. Based on the identified correspondences, we further design a new approach of multi-robot collaborative object localization to improve the object localization results by integrating multi-robot observations. Extensive experiments are conducted to evaluate our method both in robotics simulations and using physical robots, in the scenarios of connected autonomous driving, multi-robot coordination, and multi-robot collaborative object localization. The experimental results have shown that our approach obtains the state-of-the-art performance for correspondence identification with non-covisible objects under uncertainty, and also that our approach can be well integrated with our proposed state fusion algorithm for multi-robot collaborative object localization.

Acknowledgements This work was partially supported by the national science foundation CAREER award IIS-1942056 and army research laboratory, DCIST, CRA W911NF-17-2-0181.

References

Almohamad H, Duffuaa SO (1993) A linear programming approach for the weighted graph matching problem. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(5):522–525

Aragues R, Montijano E, Sagues C (2011) Consistent data association in multi-robot systems with limited communications. In: Robotics: Science and Systems

Bertoni L, Kreiss S, Alahi A (2019) MonoLoco: Monocular 3D pedestrian localization and uncertainty estimation. In: IEEE International Conference on Computer Vision

Besl PJ, McKay ND (1992) Method for registration of 3D shapes. In: Sensor Fusion, vol 1611, pp 586–606

Black J, Ellis T (2002) Multi-camera image measurement and correspondence. Measurement 32(1):61–71

Boroson ER, Ayanian N (2019) 3D Keypoint Repeatability for Heterogeneous Multi-Robot SLAM. In: IEEE International Conference on Robotics and Automation

Boyd S, Parikh N, Chu E, Peleato B, Eckstein J, et al. (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning 3(1):1–122

Brambilla M, Ferrante E, Birattari M, Dorigo M (2013) Swarm robotics: a review from the swarm engineering perspective. Swarm Intelligence 7(1):1–41

Brooks S, Gelman A, Jones G, Meng XL (2011) Handbook of markov chain monte carlo. CRC press

Buschka P, Saffiotti A, Wasik Z (2000) Fuzzy landmark-based localization for a legged robot. In: International Conference on Intelligent Robots and Systems, vol 2, pp 1205–1210

Carreras C, Walker ID (2001) Interval methods for fault-tree analysis in robotics. IEEE Transactions on Reliability 50(1):3–11

Chang HJ, Fischer T, Petit M, Zambelli M, Demiris Y (2017) Learning kinematic structure correspondences using multi-order similarities. IEEE Transactions on Pattern Analysis and Machine Intelligence (1):1–1

Chen Y, Zhu X, Gong S (2017) Person re-identification by deep learning multi-scale representations. In: IEEE International Conference on Computer Vision

Cho M, Lee J, Lee KM (2010) Reweighted random walks for graph matching. In: European Conference on Computer Vision

Chung SJ, Paranjape AA, Dames P, Shen S, Kumar V (2018) A survey on aerial swarm robotics. IEEE Transactions on Robotics 34(4):837–855

Der Kiureghian A, Ditlevsen O (2009) Aleatory or epistemic? Does it matter? Structural Safety 31(2):105–112

Ding N, Fang Y, Babbush R, Chen C, Skeel RD, Neven H (2014) Bayesian sampling using stochastic gradient thermostats. In: Advances in neural information processing systems, pp 3203–3211

Dogar M, Spielberg A, Baker S, Rus D (2019) Multi-robot grasp planning for sequential assembly operations. Autonomous Robots 43(3):649–664

Duchenne O, Bach F, Kweon IS, Ponce J (2011) A tensor-based algorithm for high-order graph matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(12):2383–2395

Engel J, Schöps T, Cremers D (2014) LSD-SLAM: Large-scale direct monocular SLAM. In: European Conference on Computer Vision

Fathian K, Khosoussi K, Lusk P, Tian Y, How JP (2019) CLEAR: A consistent lifting, rmbedding, and alignment rectification algorithm for multi-agent data association. arXiv

Frey KM, Steiner TJ, How JP (2019) Efficient constellation-based map-merging for semantic SLAM. In: IEEE International Conference on Robotics and Automation

Gao P, Guo R, Lu H, Zhang H (2020a) Regularized graph matching for correspondence identification under uncertainty in collaborative perception. Robotics: Science and Systems

Gao P, Reily B, Paul S, Zhang H (2020b) Visual reference of ambiguous objects for augmented reality-powered human-robot communication in a shared workspace. International Conference on Virtual, Augmented and Mixed Reality

Girshick R (2015) Fast R-CNN. In: IEEE international conference on computer vision, pp 1440–1448

Gojcic Z, Zhou C, Wegner JD, Wieser A (2019) The perfect match: 3D point cloud matching with smoothed densities. In: IEEE Conference on Computer Vision and Pattern Recognition

Gundavarapu NB, Srivastava D, Mitra R, Sharma A, Jain A (2019) Structured Aleatoric Uncertainty in Human Pose Estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, Workshops

He K, Gkioxari G, Dollár P, Girshick R (2017) Mask R-CNN. In: IEEE international conference on computer vision, pp 2961–2969

Hollinger GA, Englot B, Hover F, Mitra U, Sukhatme GS (2012) Uncertainty-driven view planning for underwater inspection. In: IEEE International Conference on Robotics and Automation, pp 4884–4891

Hong H, Yu H, Lee BH (2019) Regeneration of normal distributions transform for target lattice based on fusion of truncated Gaussian components. IEEE Robotics and Automation Letters 4(2):684–691

Hu J, Zhang Y, Okatani T (2019) Visualization of convolutional neural networks for monocular depth estimation. In: IEEE International Conference on Computer Vision, pp 3869–3878

Hu N, Huang Q, Thibert B, Guibas LJ (2018) Distributable consistent multi-object matching. In: IEEE Conference on Computer Vision and Pattern Recognition

Kallasi F, Rizzini DL, Caselli S (2016) Fast keypoint features from laser scanner for robot localization and mapping. IEEE Robotics and Automation Letters 1(1):176–183

Kendall A, Gal Y (2017) What uncertainties do we need in bayesian deep learning for computer vision? In: Advances in Neural Information Processing Systems

Kendall A, Badrinarayanan V, Cipolla R (2015a) Bayesian Segnet: Model uncertainty in deep convolutional encoder-

decoder architectures for scene understanding. arXiv

Kendall A, Grimes M, Cipolla R (2015b) PoseNet: A convolutional network for real-time 6-DOF camera relocalization. In: IEEE International Conference on Computer Vision.

Kendall A, Gal Y, Cipolla R (2018) Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: IEEE Conference on Computer Vision and Pattern Recognition

Korattikara Balan A, Rathod V, Murphy KP, Welling M (2015) Bayesian dark knowledge. Advances in Neural Information Processing Systems 28:3438–3446

Kraus F, Dietmayer K (2019) Uncertainty estimation in one-stage object detection. In: IEEE Intelligent Transportation Systems Conference, pp 53–60

Laina I, Rupprecht C, Belagiannis V, Tombari F, Navab N (2016) Deeper depth prediction with fully convolutional residual networks. In: International conference on 3D vision, IEEE, pp 239–248

Lampert CH, Nickisch H, Harmeling S (2014) Attribute-based classification for zero-shot visual object categorization. IEEE Transactions on Pattern Analysis and Machine Intelligence 36(3):453–465

Learning M, Rabanser S, Shchur O, Günnemann S (2015) Introduction to Tensor Decompositions and their Applications in Machine Learning. Machine Learning 98(1-2):1–5

Lee J, Cho M, Lee KM (2011) Hyper-graph matching via reweighted random walks. In: IEEE Conference on Computer Vision and Pattern Recognition

Leonardos S, Zhou X, Daniilidis K (2017) Distributed consistent data association via permutation synchronization. In: IEEE International Conference on Robotics and Automation

Leordeanu M, Hebert M (2005) A spectral technique for correspondence problems using pairwise constraints. In: IEEE International Conference on Computer Vision

Li Q, Xiong R, Vidal-Calleja T (2017) A GMM based uncertainty model for point clouds registration. Robotics and Autonomous Systems 91:349–362

Li Y, Gal Y (2017) Dropout inference in Bayesian neural networks with alpha-divergences. In: International Conference on Machine Learning

Li Y, Hernández-Lobato JM, Turner RE (2015) Stochastic expectation propagation. Advances in neural information processing systems

Maset E, Arrigoni F, Fusiello A (2017) Practical and efficient multi-view matching. In: IEEE International Conference on Computer Vision

Munkres J (1957) Algorithms for the assignment and transportation problems. Journal of the Society for Industrial and Applied Mathematics 5(1):32–38

Mur-Artal R, Montiel JMM, Tardos JD (2015) ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Transactions on Robotics 31(5):1147–1163

Neal RM (2012) Bayesian learning for neural networks, vol 118. Springer Science & Business Media

Neiswanger W, Wang C, Xing E (2013) Asymptotically exact, embarrassingly parallel MCMC. arXiv preprint arXiv:13114780

Neudecker H (1969) A note on Kronecker matrix products and matrix equation systems. SIAM Journal on Applied Mathematics 17(3):603–606

Nguyen A, Ben-Chen M, Welnicka K, Ye Y, Guibas L (2011) An optimization approach to improving collections of shape maps. In: Computer Graphics Forum, vol 30, pp 1481–1491

Nguyen Q, Gautier A, Hein M (2015) A flexible tensor block coordinate ascent scheme for hypergraph matching. In: IEEE Conference on Computer Vision and Pattern Recognition

Pachauri D, Kondor R, Singh V (2013) Solving the multi-way matching problem by permutation synchronization. In: Advances in Neural Information Processing Systems

Reily B, Reardon C, Zhang H (2020) Representing Multi-Robot Structure through Multimodal Graph Embedding for the Selection of Robot Teams. arXiv preprint arXiv:200312164

Richard MD, Lippmann RP (1991) Neural network classifiers estimate bayesian a posterior probabilities. Neural computation 3(4):461–483

Robin C, Lacroix S (2016) Multi-robot target detection and tracking: taxonomy and survey. Autonomous Robots 40(4):729–760

Senanayake M, Senthooran I, Barca JC, Chung H, Kamruzzaman J, Murshed M (2016) Search and tracking algorithms for swarms of robots: A survey. Robotics and Autonomous Systems 75:422–434

Sobreira H, Costa CM, Sousa I, Rocha L, Lima J, Farias P, Costa P, Moreira AP (2019) 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

Suh Y, Adamczewski K, Mu Lee K (2015) Subgraph matching using compactness prior for robust feature correspondence. In: IEEE Conference on Computer Vision and Pattern Recognition

Szeliski R (2010) Computer vision: Algorithms and applications. Springer Science & Business Media

Thrun S (2002) Probabilistic robotics. Communications of the ACM 45(3):52–57

Tian Y, Liu K, Ok K, Tran L, Allen D, Roy N, How JP (2019) Search and rescue under the forest canopy using multiple UAVs. arXiv

Tron R, Zhou X, Esteves C, Daniilidis K (2017) Fast multi-image matching via density-based clustering. In: IEEE International Conference on Computer Vision

Wang R, Yan J, Yang X (2019) Learning combinatorial embedding networks for deep graph matching. In: IEEE International Conference on Computer Vision

Wei S, Yu D, Guo CL, Dan L, Shu WW (2018) Survey of connected automated vehicle perception mode: from autonomy to interaction. IET Intelligent Transport Systems 13(3):495–505

Yan J, Ren Z, Zha H, Chu S (2016) A constrained clustering based approach for matching a collection of feature sets. In: International Conference on Pattern Recognition

Yan Z, Jouandeau N, Cherif AA (2013) A survey and analysis of multi-robot coordination. International Journal of Advanced Robotic System 10(12):399

Zhang Q, Pless R (2004) Extrinsic calibration of a camera and laser range finder (improves camera calibration). In: IEEE/RSJ International Conference on Intelligent Robots and Systems

Zhang R, Wang W (2016) An MCMC-based prior sub-hypergraph matching in presence of outliers. In: International Conference on Pattern Recognition

Zhao R, Oyang W, Wang X (2016) Person re-identification by saliency learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(2):356–370

Zhao Y, Shen X, Jin Z, Lu H, Hua Xs (2019) Attribute-driven feature disentangling and temporal aggregation for video person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition

Zhou X, Zhu M, Daniilidis K (2015) Multi-image matching via fast alternating minimization. In: IEEE International Conference on Computer Vision