2020 IEEE International Conference on Robotics and Automation (ICRA)
31 May - 31 August, 2020. Paris, France

Correspondence Identification in Collaborative Robot Perception through Maximin Hypergraph Matching

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

***Abstract*— Correspondence identification is an essential problem for collaborative multi-robot perception, with the objective of deciding the correspondence of objects that are observed in the field of view of each robot. In this paper, we introduce a novel maximin hypergraph matching approach that formulates correspondence identification as a hypergraph matching problem. The proposed approach incorporates both spatial relationships and appearance features of objects to improve representation capabilities. It also integrates the maximin theorem to optimize the worst-case scenario in order to address distractions caused by non-covisible objects. In addition, we design an optimization algorithm to address the formulated non-convex non-continuous optimization problem. We evaluate our approach and compare it with seven previous techniques in two application scenarios, including multi-robot coordination on real robots and connected autonomous driving in simulations. Experimental results have validated the effectiveness of our approach in identifying object correspondence from partially overlapped views in collaborative perception, and have shown that the proposed maximin hypergraph matching approach outperforms previous techniques and obtains state-of-the-art performance.**

I. Introduction

Multi-robot systems have been attracting an increasing attention over the past years, because of their advantages of parallelism, reliability, and flexibility to efficiently perform collaborative tasks [1], [2], [3]. Collaborative perception is a critical capability required by multi-robot systems to collaboratively understand the environment for shared situational awareness and effective teamwork. Multi-robot collaborative perception is widely applied to a range of real-world applications, including search and rescue [4], [5], [6], homeland security [7], manufacture [8] and connected autonomous driving [9], in which multiple robots collaboratively perceive and operate together as a team [10], [11].

To enable collaborative perception, correspondence identification must be addressed, with the objective of determining the correspondence between objects observed in the field of view of each robot in a multi-robot system [12], [13], [14]. Here, we utilize the term *objects* to broadly refer to robots, humans, and other entities of interest in the environment. Figure 1 depicts a scenario of correspondence identification in collaborative multi-robot perception: Before the ground robot and the aerial robot communicate with or manipulate another object, each robot must identify the object within its own field of view, so that the pair of robots correctly refer to the same object.

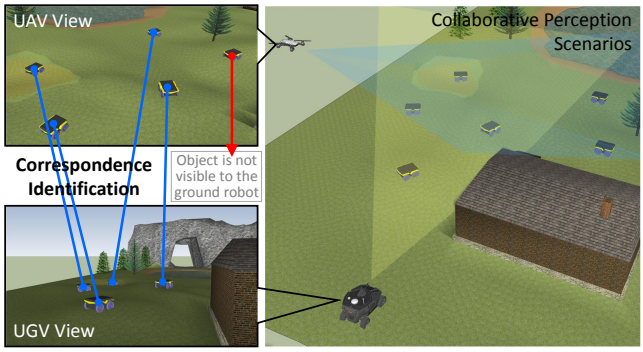


Fig. 1: A motivating example of correspondence identification in collaborative multi-robot perception. Before an aerial vehicle and a ground robot collaboratively track an object, the pair of robots must correctly refer to the same object in their own fields of view.

Correspondence identification is difficult to solve in collaborative multi-robot perception due to several challenges. First, objects in the field of view of one robot may not be observed by other robots, because of the limited fields of view of the robots and occlusion by other objects. Second, the same object can look different when they are observed from different viewpoints by a pair of robots. Illumination angles and sensor noise may cause the same problem. Third, object appearances may look similar and even identical, for example, when robots with the same type are used in a team. Although several approaches were previously implemented in various applications, including re-identification for individual object matching [15], keypoint-based or dense point association [16], and graph or hypergraph-based matching [17], they cannot well address the challenges of correspondence identification from partially overlapped robot views.

In this paper, we propose a principled approach to address object correspondence identification in multi-robot collaborative perception. We first develop a hypergraph representation that integrates appearance cues and spatial relationships of observed objects to improve the expressiveness of the representation. Then, we formulate correspondence identification as a hypergraph matching problem. Inspired by the maximin theorem, we introduce a novel *maximin* hypergraph matching approach that optimizes the worst-case scenario to identify

*This work was partially supported by NSF IIS-1942056, IIS-1849348, CNS-1823245, and unrestricted research fund offered by Toyota Motor North America.

[1]Peng Gao, Ziling Zhang and Hao Zhang are with the Human-Centered Robotics Laboratory at the Colorado School of Mines, Golden, CO 80401. {gaopeng, zilingzhang, hzhang}@mines.edu

[2]Rui Guo and Hongsheng Lu are with the Toyota Motor North America, Mountain View, CA 94043. {rgou, hlu}@us.toyota-itc.com

object correspondence from partially overlapped robot views. We evaluate our approach in two representative application scenarios, including multi-robot coordination and connected autonomous driving. Experimental results have shown state-of-the-art performance of our maximin hypergraph matching approach to identify object correspondence for collaborative perception.

The novelty of this paper is twofold. First, we formulate correspondence identification as a novel maximin hypergraph matching problem, which is able to integrate objects' spatial relationships and appearance cues to improve representation expressiveness, and utilize the maximin theorem to optimize the worst-case scenario to better identify the correspondence given partially overlapped observations. Second, because the optimization problem in our formulation is non-convex and non-continuous, we implement a new algorithm to effectively solve the formulated optimization problem.

II. Related Work

A. Multi-Robot Collaborative Perception

Multi-robot collaborative perception has been widely studied. Views from multiple agents were merged using iterative closest points (ICP) in connected driving applications [10]. Spatiotemporal perceptual data from multiple vehicles was fused using the extended Kalman filter (EKF) to perceive complex road surfaces [18]. Localization accuracy was improved by factor graphs to fuse radar data from connected vehicles [19]. Recently, collaborative perception to monitor objects using a team of robots has attracted an increasing attention. For example, multiple aerial robots were used to collaboratively track people [20]; ground vehicles employed collaborative perception to improve prediction of occluded vehicles [11]; underwater robots collaboratively tracked a target to perform underwater multi-robot convoying [21].

B. Correspondence Identification

To enable collaborative perception by multiple robots, correspondence identification is a fundamental challenge. Existing techniques for correspondence identification can be generally categorized into three groups: point-based association, re-identification, and graph/hypergraph-based matching.

Point-based association is widely used in reconstruction, such as matching adjacent frames in simultaneous localization and mapping (SLAM). Dense point association was implemented to match between most of points in pair of frames, e.g., based upon ICP [22] and random sample consensus (RANSAC) [23]. Keypoint-based association extract keypoints from frames and match associate these keypoints, e.g., based on SIFT [24] or ORB [25] keypoints in SLAM. *Re-identification* methods identify correspondence of individual objects with changing appearance or viewing angles [26]. Re-identification is often performed by matching visual features [27], object attributes [28] or spatial layout [15]

Point-based association typically assumes that points satisfy a transformation as a constraint, which cannot be applied to identify correspondence of dynamic and independent objects. Point-based techniques also cannot incorporate region-based appearance cues. Re-identification techniques focus on identifying correspondence of an individual object, and are unable to incorporate relationships of multiple objects for matching.

C. Graph and Hypergraph Matching

Graph and hypergraph matching provides a promising paradigm to match points and objects with unstructured relationships. Pairwise graphs have segments as edges, and hypergraphs use tuples (such as triangles) as edges [17].

For graph matching, [29] identified point correspondence by exploring principal eigenvector of the affinity matrix. [30] searched correspondence through factorizing a large affinity matrix into smaller matrices that encode local relationships. [31] addressed the non-convex point association problem using a random walk algorithm. [32] developed a path following method to solve the optimization. Compactness prior was used to improve matching [33].

It is widely recognized that hypergraph matching is more robust to geometric variations and noise by integrating high-order relationships. [34] designed tensor-based high-order constraints to encode the similarity of high-order hyperedges. [17] designed a tensor-based reweighted random walk algorithm with reweighting jumps. [35] formulated the problem in a lower dimension by factorization. [36] proposed a tensor block coordinate ascent algorithm as a solver for hypergraph association. [37] optimized in the discrete domain by linear assignment approximation.

Almost all existing graph and hypergraph matching methods focus on identifying correspondence of points and do not consider appearances of objects. [38] fused relationships and appearance cues in a linear combination to match body joints of humans and humanoid robots. This approach assumes that body joints follow fixed kinematic structures (i.e., body skeleton) as prior knowledge and all joints can be well observed. Because of object independency, robot autonomy, and view occlusion in multi-robot collaborative perception, existing graph and hypergraph matching methods are not directly applicable.

III. The Proposed Approach

A. Formulating Correspondence Identification as a Hypergraph Matching Problem

Given the observations obtained by a robot (e.g., from a color-depth camera), the observed environment is represented as a hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ with $\boldsymbol{v}_i$ denoting the 3D position of the i-th object instance, and n is the number of object instances observed in the environment. We model the high-order spatial relationships of the object instances in $\mathcal{V}$ using a set of hyperedges $\mathcal{E} = \{\boldsymbol{e}_{i,j,k}\}$, with each hyperedge $\boldsymbol{e}_{i,j,k} = [\theta_i, \theta_j, \theta_k], i, j, k = 1, 2, \ldots, n, i \neq j \neq k$ defined to represent three angles of a triangular relationship constructed by the i-th, j-th, and k-th object instances in $\mathcal{V}$. Third order spatial relationships are

robust to scale change since angles of triangular relationships are invariant to scale change.

In collaborative perception, we assume that a pair of robots obtain partially overlapped observations of the same environment with *covisible* objects (i.e., instances observed by both robots) and *non-covisible* objects (i.e., instances observed by one robot only). We denote the observations obtained by the pair of robots as hypergraphs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$, respectively, which may include different numbers of objects (i.e., n' can be different from n). Then, we can compute the affinity tensor $\mathbf{T} = \{t_{ii',jj',kk'}\} \in \mathbb{R}^{nn' \times nn' \times nn'}$ between $\mathcal{G}$ and $\mathcal{G}'$, with each $t_{ii',jj',kk'}$ computed from a pair of hyperedges $\boldsymbol{e}_{ijk} \in \mathcal{E}$ and $\boldsymbol{e}'_{i'j'k'} \in \mathcal{E}'$. Due to third order spatial relationsihps are robustness to scale variations, an angular similarity function [38], [17] is adopted to compute the similarity of hyperedges $\boldsymbol{e}_{i,j,k}$ and $\boldsymbol{e}'_{i',j',k'}$ as:

$$t_{ii',jj',kk'} = \exp\left(-\frac{1}{\sigma} \sum_{p \in i,j,k; p' \in i',j',k'} |\cos(\theta_p) - \cos(\theta_{p'})|\right) \quad (1)$$

where θ_p denotes the angle with $\boldsymbol{v}_p$ as the vertex, and σ is used to control the magnitude of input of exp function. We set $\sigma = 0.5$ empirically.

Then, correspondence identification of objects observed by two robots in collaborative perception can be formulated as a hypergraph matching task by solving the following problem:

$$\mathbf{X}^* = \arg\max_{\mathbf{X}} \sum_{ii'=1}^{nn'} \sum_{jj'=1}^{nn'} \sum_{kk'=1}^{nn'} t_{ii',jj',kk'} x_{ii'} x_{jj'} x_{kk'}$$
$$\text{s.t.} \quad \mathbf{X}\mathbf{1}_{n' \times 1} \leq \mathbf{1}_{n \times 1}, \mathbf{X}^\top \mathbf{1}_{n \times 1} \leq \mathbf{1}_{n' \times 1} \quad (2)$$

The objective function in Eq. (2) models the accumulative similarity among all hyperedges in $\mathcal{G}$ and $\mathcal{G}'$ (parameterized by $\mathbf{X}$). Eq. (2) aims to find the optimal $\mathbf{X}*$ that maximizes the accumulative similarity. We can re-write Eq. (2) into a matrix form as:

$$\mathbf{X}^* = \arg\max_{\mathbf{X}} \ \mathbf{T} \otimes_1 \mathbf{x} \otimes_2 \mathbf{x} \otimes_3 \mathbf{x} \quad (3)$$
$$\text{s.t.} \quad \mathbf{X}\mathbf{1}_{n' \times 1} \leq \mathbf{1}_{n \times 1}, \mathbf{X}^\top \mathbf{1}_{n \times 1} \leq \mathbf{1}_{n' \times 1}$$

where $\mathbf{x} = \{x_{ii'}\} \in \{0,1\}^{nn'}$ is the vectorized form of correspondence matrix $\mathbf{X} \in \{0,1\}^{n \times n'}$ with $x_{ii'} = 1$ denoting the i-th node in $\mathcal{V}$ and the i'-th node in $\mathcal{V}'$ are matched, $\mathbf{1}$ is an all-ones vector, $\otimes$ is a tensor product, and $\otimes_j$ means multiplication between $\mathbf{X}$ and the mode-$j, j = 1, 2, 3$, matricization of $\mathbf{T}$ [39].

B. Maximin Hypergraph Matching

We propose a novel maximin matching approach to incorporate both appearance cues of object instances and their high-order spatial similarity in a principled maximin optimization framework that optimizes the worst case.

1) Integrating Appearances in Hypergraph: Only considering hyperedge similarities for correspondence identification (e.g., in Eq. (2)) often results in incorrect matches when a pair of hypergraphs exhibit deformations caused by differences in viewing perspectives. Different from point-to-point matching, objects themselves can provide informative cues that can be used for correspondence identification in our case. Thus, to increase the expressiveness of our hypergraph-based representation, we integrate appearance cues of the objects into the hypergraph matching approach that takes into account of similarities of both hyperedges and the node appearances associated with these hyperedges.

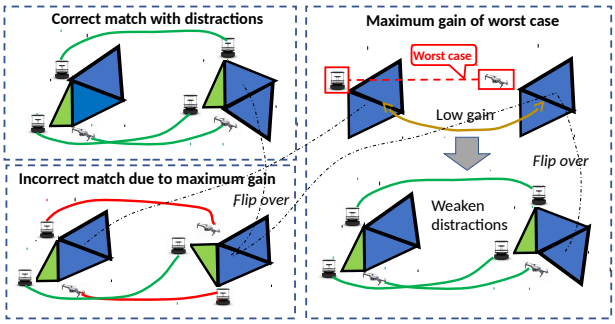


Fig. 2: Illustration of the maximin matching. Green/red lines denote correct/incorrect matches and blue hyperedges denote the distraction introduced by non-covisible objects. The *top-left figure* shows the correct match between two hypergraphs with distractions. The *bottom-left figure* illustrates that maximizing the average gain can lead to an incorrect match. The *right figure* shows our proposed maximin approach is robust to distraction and obtain the correct match by optimizing the worst-case scenario.

Formally, for each node $\boldsymbol{v}$, a feature vector $\mathbf{c} \in \mathbb{R}^d$ is computed from the object associated with $\boldsymbol{v}$, where d denotes the dimensionality of the feature vector. The feature vector $\mathbf{c}$ can include color, shape, and texture features, or a concatenation of them. Given $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the feature set extracted from all objects is denoted as $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_n\}$. Then, given a pair of hypergraphs $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$, and the respective feature sets $\mathcal{C}$ and $\mathcal{C}'$, we compute the appearance similarity vector $\mathbf{b} = \{b_{ii'}\} \in \mathbb{R}^{nn'}$, where each element $b_{ii'}$ represents the similarity between the i-th object encoded by $\boldsymbol{v}_i \in \mathcal{V}$ and the i'-th object encoded by $\boldsymbol{v}'_i \in \mathcal{V}'$. The similarity $b_{ii'}$ can be computed by a distance function over the objects' feature vectors, for example, based on the cosine distance between $\mathbf{c}_i$ and $\mathbf{c}'_{i'}$ such as $b_{ii'} = \frac{\mathbf{c}_i \cdot \mathbf{c}'_i}{\|\mathbf{c}_i\| \|\mathbf{c}'_i\|}$.

2) Maximin Optimization for Hypergraph Matching: The existence of non-covisible objects introduces distractions (the spatial relationships constructed with non-covisible objects distract the correct match of co-visible objects) into hypergraph matching and significantly increases the possibility of incorrect correspondences, as depicted in Figure 2. In addition, before object correspondences are identified, it is impractical to identify non-covisible objects individually, due to object dynamics and similar appearance (as shown in Figure 1).

To address the challenge of non-covisible objects, we introduce the principled *maximin hypergraph matching* approach to integrate spatial hyperedge similarities and node

appearance similarities in a unified optimization framework. Our method is inspired by the maximin theorem[1] that maximizes the possible gain for the worst-case (i.e., minimum gain) scenario. Without optimizing the overall gain of all possible cases, the *maximin* value is the highest gain that an approach can be sure to get. In Figure 2, the worst case is when the similarity of object appearances is the smallest and traditional methods which generally maximize the spatial and appearance similarity to find the matches [38]. However, due to the distractions introduced by non-covisible objects, maximum similarity often leads to incorrect matches (see the *Bottomleft Figure* in Figure 2). Our proposed maximin approach to optimize the worst case, which can address distractions caused by non-covisible objects and improve the robustness against matching with non-covisible objects (An illustrative example is provided in Figure 2).

Formally, we formulate correspondence identification with non-covisible objects as the novel maximin optimization problem, which maximizes the similarity under the worst case to decide correspondences:

$$\begin{aligned}\mathbf{X}^* = \arg\max_{\mathbf{X}} & \sum_{ii'=1}^{nn'}\sum_{jj'=1}^{nn'}\sum_{kk'=1}^{nn'} \\ & t_{ii',jj',kk'}x_{ii'}x_{jj'}x_{kk'}\min\{b_{ii'},b_{jj'},b_{kk'}\} \\ \text{s.t.} \quad & \mathbf{X}\mathbf{1}_{n'\times 1}\leq \mathbf{1}_{n\times 1}, \mathbf{X}^\top \mathbf{1}_{n\times 1}\leq \mathbf{1}_{n'\times 1} \end{aligned} \tag{4}$$

After solving the optimization problem in Eq. (4) based on Algorithm 1, we obtain an optimal solution $\mathbf{X}^*$ that includes the correspondence of $\mathcal{V}$ and $\mathcal{V}'$ (i.e., object correspondence). Our maximin hypergraph matching approach has several advantages.

By integrating hyperedge and node similarities, our approach improves the expressiveness of representation and matching performance. By formulating hypergraph matching under the maximin optimization theorem, our approach is robust to scenarios with non-covisible objects.

C. Optimization Algorithm

Since our proposed hypergraph matching formulation is a non-convex non-continuous optimization problem with a non-smooth minimize operator, we design a new heuristic optimization algorithm based on random walk with the reweighted jump technique [17] for the proposed optimization problem in in Eq. (4). Our optimization algorithm is presented in Algorithm 1.

In Step 2, we calculate the spatial similarity under the worst case $\mathbf{T}' = \{t'_{ii',jj',kk'}\} \in \mathbb{R}^{nn'\times nn'\times nn'}$ as follows,

$$t'_{ii',jj',kk'} = t_{ii',jj',kk'}\min\{b_{ii'},b_{jj'},b_{kk'}\} \tag{5}$$

and we convert the tensor $\mathbf{T}'$ to a stochastic form $\mathbf{P} = \{p_{ii',jj',kk'}\} \in \mathbb{R}^{nn'\times nn'\times nn'}$ as follows:

$$\mathbf{P} = \mathbf{T}' / \max_i \sum_{j,k} \mathbf{T}'_{i,j,k} \tag{6}$$

[1]When dealing with losses, the maxmin theorem [40] is also referred to as "minimax" that minimizes the maximum loss for a worst-case scenario.

Algorithm 1: The proposed algorithm to solve the formulated non-convex optimization problem in Eq. (4).

Input : $\mathbf{T} \in \mathbb{R}^{nn'\times nn'\times nn'}$ and $\mathbf{b} \in \mathbb{R}^{nn'}$
Output: $\mathbf{X} = \in \{0,1\}^{n\times n'}$

1: Initialize the correspondence matrix $\mathbf{X}$
2: Compute $\mathbf{P}$ according to Eq. (5) and Eq. (6)
3: **while** *not converge* **do**
4: Compute the jump vector $\mathbf{z}$ by Eq. (7)
5: Normalize $\mathbf{z}$ using the bistochastic normalization
6: Update $\mathbf{X}$ with reweighted jump by Eq. (8)
7: **end**
8: Discretize $\mathbf{X}$ using the Hungarian algorithm
9: **return** $\mathbf{X}$

Eq.(6) aims to normalize the original tensor without losing relative affinity by dividing the maximum through mode-1 matricization. Since $\mathbf{T}$ is supersymmetric, the matricization of $\mathbf{T}$ in different modes are equivalent. Then in Step 4, in order to jump out local optima, inspired by the PageRank algorithm [41], [17], we design a reweighting jump vector $\mathbf{z} \in \mathbb{R}^{n\times n'}$ as:

$$\mathbf{z}^r = \exp(\mathbf{x}^r \circ \mathbf{b} / \max(\mathbf{x}^r \circ \mathbf{b})) \tag{7}$$

where $\circ$ denotes the entrywise product and $\mathbf{x} \in \mathbb{R}^{nn'}$ is the vectorized form of the input matrix $\mathbf{X} \in \mathbb{R}^{n\times n'}$. The node appearance similarity $\mathbf{b}$ is used to guide the jump toward a direction that can better match similar objects. r denotes the $r-$th iteration.

Step 5 employs a bistochastic normalization to normalize each row and column in $\mathbf{z}$, thus enforcing the one-to-one correspondence. Then, in Step 6, to facilitate $\mathbf{X}$ to jump out of local optima, $\mathbf{X}$ is updated by:

$$\mathbf{x}^{r+1} = \alpha \mathbf{P} \otimes_2 \mathbf{x}^r \otimes_3 \mathbf{x}^r + (1-\alpha)\mathbf{z}^r \tag{8}$$

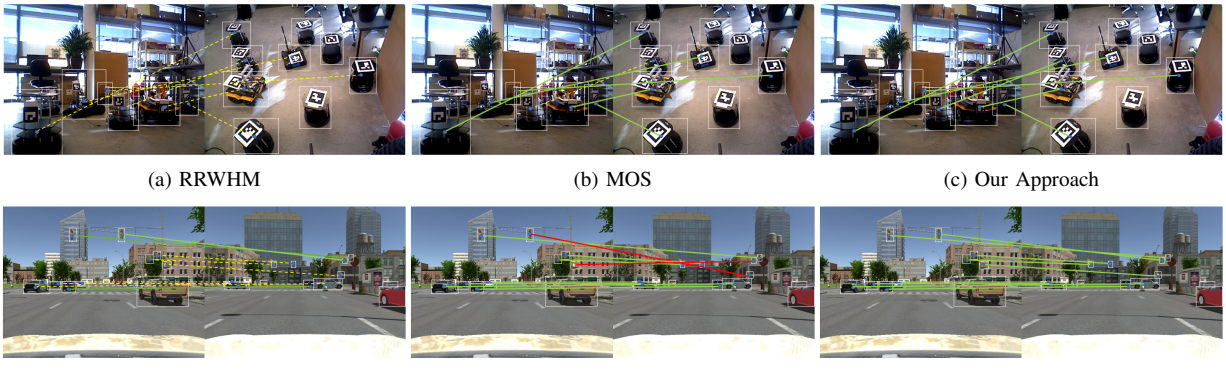
where α is a hyper-parameter that controls the update rate, and $\alpha = 0.3$ in the following experiments.

In Step 8, after algorithm convergence, we discretize $\mathbf{X}$ to obtain a binary matrix $\mathbf{X} \in \{0,1\}^{n\times n'}$ using the Hungarian algorithm.

Complexity. The *space complexity* of our maximin formulation in Eq. (4) is $O(n^6)$, dominated by the size of $\mathbf{T}$. When nearest neighborhoods are applied to compute matches locally [36], the space complexity becomes $O(n^2k)$, where k is the number of nearest neighborhoods. In this work, we set $k = n^2$, resulting in the complexity $O(n^4)$. The *time complexity* of each iteration in Algorithm 1 is $O(n^4)$, dominated by Eq. (8) to access $\mathbf{P}$ that is computed from $\mathbf{T}$ with $O(n^4)$ elements.

IV. Experiment

Extensive experiments are conducted to evaluate our maximin hypergraph matching method for object correspondence identification in two scenarios: multi-robot coordination



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

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

Fig. 3: Qualitative experimental results over the MRC (first row) and CAD (second row) datasets and comparisons with the RRWHM and MOS approaches. Green/red solid lines denote correct/incorrect correspondence; and yellow dashed lines denote missing correspondence (i.e., false negatives).

(MRC) and connected autonomous driving (CAD) as shown in Figure 3. In MRC, a team of robots is observed with partially overlapped views using a pair of 3D structured-light cameras from the side view by a mobile robot and an overhead view by an aerial robot. In CAD, a connected autonomous driving simulator is used to simulate collaborative perception behaviors when two connected vehicles observe the same intersection with partially overlapped views from different perspectives. A dataset is collected from each of two scenarios, which contains 50 pairs of color-depth images from two robots with different perspective of robots and object configurations in the environments. Each data instance includes covisible and non-covisible objects (caused by occlusion and robot's limited field of view), and includes objects with similar or identical appearances. Both datasets include 3D object position, 2D bounding boxes, object appearance features (including visual features including HOG, color histograms [42] and attribute feature [43]) and object correspondence ground truth. For example, QR code labels are used in MRC to obtain the correspondence ground truth, and the CAD simulator directly provides the ground truth.

We utilize accuracy, precision and recall as the standard metrics for performance evaluation, following [38], [44]. *Accuracy* is defined as the number of correct matches over the total number of co-visible objects. From the perspective of object retrieval, *precision* is defined as the fraction of co-visible objects over all retrieved objects, and *recall* is defined as the ratio of retrieved co-visible objects over all of the co-visible objects.

Furthermore, we compare the proposed approach with seven previous correspondence identification methods. They are two pairwise graph matching techniques, including (**SM** [29] and **RRWM** [31]) which use 2D affinity matrix to encode similarities of pairwise edges, four hypergraph matching methods, including (**TM** [34], **HGM** [35], **BCAGM** [36], and **RRWHM** [17]) which use affinity tensor to represent third order spatial similarities encoded by hyperedges, and one approach based upon Multi-Order Similarities (**MOS**) [38] which is based on RRWHM but considered multi-order similarity to match. All these methods are based on maximum gains without considering the worst case.

A. Results on the MRC Dataset

We perform experiments on MRC to evaluate our approach in a multi-robot coordination scenario. Most object instances in MRC are robots belonging to the same category with similar appearance. The overhead view can well observe the objects, but the side view contains strong occlusions.

The quantitative correspondence identification results obtained by our maximin hypergraph matching method are presented in Table I, along with comparisons with seven popular graph/hypergraph-based correspondence identification techniques. It is observed that graph matching methods (SM and RRWM) perform badly, due to different perspectives that dramatically change the spatial distance between objects in the image space. The hypergraph matching methods (HGM, TM, BCAGM, and RRWHM) obtain improved performance when they use high-order spatial relationships of the objects. When linearly combining multi-order similarities (e.g., spatial relationship and appearance), MOS further improves correspondence results. Our maximin approach obtains the best performance on MRC, and outperforms MOS, due to our approach's capability of dealing with non-covisible objects based on the maximin theorem.

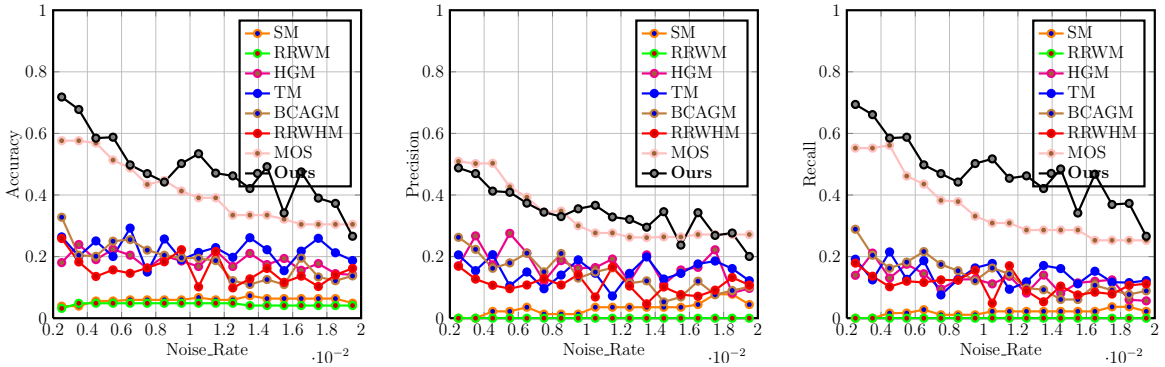
To visualize object correspondence, the qualitative experimental results of correspondence identification on a representative data instance in MRC are presented in Figure 3. Results obtained by the other two best performing methods (MOS and RRWHM) are also compared in the figure. It is observed that RRWHM cannot well identify correspondence of the objects causing a large number of false negatives in this data instance. MOS obtains improved performance with no incorrect correspondence and only one missed match. For this data instance, our maximin approach obtains the best results on object correspondence identification.

B. Results on the CAD Dataset

We also perform experiments using the dataset collected from autonomous driving simulations. The environment in-

TABLE I: Quantitative experimental results on the MRC and CAD datasets. The results are presented as $mean \pm covariance$ (%), which are computed by executing these methods four times over the datasets using different initializations.

Method	Results on the MRC dataset			Results on the CAD dataset		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
SM [29]	16.35± 0	16.67 ± 0	7.14 ± 0	3.93 ± 0	0 ± 0	0 ± 0
RRWM [31]	13.49 ± 0	8.33 ± 0	4.76 ± 0	3.19 ± 0	0 ± 0	0 ± 0
HGM [35]	39.01 ± 3.79	25.00 ± 6.70	19.05 ± 3.46	18.73 ± 6.12	10.67 ± 9.61	8.00 ± 6.02
TM [34]	44.56 ± 9.89	50.83 ± 11.67	35.87 ± 10.71	18.49 ± 2.41	6.67 ± 2.44	7.00 ± 2.30
BCAGM [36]	54.44 ± 2.60	52.78 ± 4.55	49.01 ± 2.10	25.63 ± 4.80	17.22 ± 3.69	14.22 ± 5.37
RRWHM [17]	58.77 ± 7.40	63.89 ± 6.97	53.61 ± 6.65	18.15 ± 9.98	7.78 ± 11.49	7.56 ± 12.05
MOS [38]	88.73 ±2.81	89.29 ± 2.81	79.84 ± 3.46	57.63 ± 0	**50.93 ± 0**	52.56 ± 0
Our Approach	**94.17 ± 4.73**	**91.67 ± 5.09**	**82.90 ± 4.55**	**71.13 ± 6.89**	48.53 ± 3.87	**68.72 ±6.89**



(a) Accuracy (b) Precision (c) Recall

Fig. 4: Performance analysis of the previous and our methods with respect to spatial sensing noises in the CAD scenarios. The noise model $noise = depth^2 \cdot noise\_rate$ is added to depth sensors installed on simulated vehicles in the simulation.

cludes a variety of object categories, such as pedestrians, traffic lights, road signs, and various vehicles. Both vehicles' view contains strong occlusions.

The quantitative experimental results obtained from our approach over the CAD dataset are shown in Table I, which also includes comparisons with previous methods. It is observed that our maximin approach obtains significant improvements (i.e., more than 13%) on accuracy and recall, while obtaining slightly worse precision than MOS. Our approach and MOS still obtain superior performance over other techniques due to the integration of spatial relationships and appearance cues. The qualitative results of object correspondence in a representative situation are demonstrated in Figure 3, including visual comparisons of our approach with RRWHM and MOS. It is observed that RRWHM correctly matches two pairs of objects but fails to identify correspondence of the other objects in this situation. MOS is able to identify correspondence of all objects, but several matches are not correct. Our maximin hypergraph matching approach identifies the correspondence correctly in this situation.

In the experiment, we analyze the robustness of correspondence identification methods to sensing noise. We add a noise generation model to the depth sensor installed on the simulated vehicles. The model creates $noise = depth^2 \times noise\_rate$ at a given depth, which is similar to the errors obtained from stereo vision [45] and monocular depth estimation [46]. For example, the added noise is 1 meter at the depth of 10 meters when $noise\_rate = 0.1$. In Figure 4, we show performance variations of correspondence identification methods with respect to different noise-rate values. It is observed that, with the increase of the noise rate, the performance of all the hypergraph matching approaches gradually decreases with small fluctuations. The accuracy and recall curves of hypergraph matching methods show similar trends and values, since they are strongly dependent on true positives (i.e., correct matches) when the number of true negatives and false positives is small. Moreover, it is observed that our approach and MOS greatly outperform other methods under noise, and our approach obtains the best performance in most cases.

V. Conclusion

We propose a novel maximin hypergraph matching approach that formulates object correspondence identification as a hypergraph matching problem. The proposed approach integrates both spatial object relationships and appearance cues to improve representation expressiveness, and adopts the maximin theorem to optimize worst-case scenarios in order to address distractions caused by non-covisible objects. A new optimization algorithm is designed to solve the formulated non-convex maximin optimization problem. We evaluate our method in two application scenarios, including multi-robot perception of physical robots and connected autonomous driving in simulations. Experimental results have shown that our approach well identifies object correspondence from partially overlapped perspectives in collaborative perception, and obtains state-of-the-art performance of correspondence identification.

REFERENCES

[1] Z. Yan, N. Jouandeau, and A. A. Cherif, "A survey and analysis of multi-robot coordination," *International Journal of Advanced Robotic Systems*, vol. 10, no. 12, p. 399, 2013.

[2] S.-J. Chung, A. A. Paranjape, P. Dames, S. Shen, and V. Kumar, "A survey on aerial swarm robotics," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 837–855, 2018.

[3] M. Brambilla, E. Ferrante, M. Birattari, and M. Dorigo, "Swarm robotics: A review from the swarm engineering perspective," *Swarm Intelligence*, vol. 7, no. 1, pp. 1–41, 2013.

[4] C. Amato, "Decision-making under uncertainty in multi-agent and multi-robot systems: Planning and learning," in *International Joint Conferences on Artificial Intelligence*, 2018.

[5] M. Senanayake, I. Senthooran, J. C. Barca, H. Chung, J. Kamruzzaman, and M. Murshed, "Search and tracking algorithms for swarms of robots: A survey," *Robotics and Autonomous Systems*, vol. 75, pp. 422–434, 2016.

[6] C. Robin and S. Lacroix, "Multi-robot target detection and tracking: Taxonomy and survey," *Autonomous Robots*, vol. 40, no. 4, pp. 729–760, 2016.

[7] B. Kehoe, S. Patil, P. Abbeel, and K. Goldberg, "A survey of research on cloud robotics and automation," *IEEE Transactions on automation science and engineering*, vol. 12, no. 2, pp. 398–409, 2015.

[8] M. Dogar, A. Spielberg, S. Baker, and D. Rus, "Multi-robot grasp planning for sequential assembly operations," *Autonomous Robots*, vol. 43, no. 3, pp. 649–664, 2019.

[9] S. Wei, D. Yu, C. L. Guo, L. Dan, and W. W. Shu, "Survey of connected automated vehicle perception mode from autonomy to interaction," *IET Intelligent Transport Systems*, vol. 13, no. 3, pp. 495–505, 2018.

[10] S.-W. Kim, B. Qin, Z. J. Chong, X. Shen, W. Liu, M. H. Ang, E. Frazzoli, and D. Rus, "Multivehicle cooperative driving using cooperative perception: Design and experimental validation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 663–680, 2015.

[11] R. Yee, E. Chan, B. Cheng, and G. Bansal, "Collaborative perception for automated vehicles leveraging vehicle-to-vehicle communications," in *IEEE Intelligent Vehicles Symposium*, 2018.

[12] G. Ozbilgin, U. Ozguner, O. Altintas, H. Kremo, and J. Maroli, "Evaluating the requirements of communicating vehicles in collaborative automated driving," in *IEEE Intelligent Vehicles Symposium*, 2016.

[13] N. E. Özkucur, B. Kurt, and H. L. Akın, "A collaborative multi-robot localization method without robot identification," in *Robot Soccer World Cup*, 2008, pp. 189–199.

[14] A. Ravankar, A. Ravankar, Y. Kobayashi, and T. Emaru, "Symbiotic navigation in multi-robot systems with remote obstacle knowledge sharing," *Sensors*, vol. 17, no. 7, p. 1581, 2017.

[15] W.-S. Zheng, X. Li, T. Xiang, S. Liao, J. Lai, and S. Gong, "Partial person re-identification," in *IEEE International Conference on Computer Vision*, 2015.

[16] G. Georgakis, S. Karanam, Z. Wu, J. Ernst, and J. Košecká, "End-to-end learning of keypoint detector and descriptor for pose invariant 3D matching," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[17] J. Lee, M. Cho, and K. M. Lee, "Hyper-graph matching via reweighted random walks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[18] A. Rauch, F. Klanner, R. Rasshofer, and K. Dietmayer, "Car2x-based perception in a high-level fusion architecture for cooperative perception systems," in *IEEE Intelligent Vehicles Symposium*, 2012.

[19] D. Gulati, V. Aravantinos, N. Somani, and A. Knoll, "Robust vehicle infrastructure cooperative localization in presence of clutter," in *International Conference on Information Fusion*, 2018.

[20] E. Price, G. Lawless, R. Ludwig, I. Martinovic, H. Bulthoff, M. J. Black, and A. Ahmad, "Deep neural network-based cooperative visual tracking through multiple micro aerial vehicles," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3193–3200, 2018.

[21] F. Shkurti, W.-D. Chang, P. Henderson, M. J. Islam, J. C. G. Higuera, J. Li, T. Manderson, A. Xu, G. Dudek, and J. Sattar, "Underwater multi-robot convoying using visual tracking by detection," in *IEEE International Conference on Intelligent Robots and Systems*, 2017.

[22] H. Sobreira, C. M. Costa, I. Sousa, L. Rocha, J. Lima, P. Farias, P. Costa, and A. P. Moreira, "Map-matching algorithms for robot self-localization: A comparison between perfect match, iterative closest point and normal distributions transform," *Journal of Intelligent and Robotic Systems*, vol. 93, no. 3-4, pp. 533–546, 2019.

[23] G. H. Lee, F. Fraundorfer, and M. Pollefeys, "RS-SLAM: RANSAC sampling for visual FastSLAM," in *IEEE International Conference on Intelligent Robots and Systems*, 2011.

[24] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: large-scale direct monocular SLAM," in *European Conference on Computer Vision*, 2014.

[25] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[26] R. Zhao, W. Oyang, and X. Wang, "Person re-identification by saliency learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 2, pp. 356–370, 2016.

[27] M. Cristani and V. Murino, "Person re-identification," in *Academic Press Library in Signal Processing*, 2018, vol. 62, pp. 365–394.

[28] R. Zhao, W. Oyang, and X. Wang, "Person re-identification by saliency learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 2, pp. 356–370, 2017.

[29] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *IEEE International Conference on Computer Vision*, 2005.

[30] F. Zhou and F. De la Torre, "Factorized graph matching," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 9, pp. 1774–1789, 2016.

[31] M. Cho, J. Lee, and K. M. Lee, "Reweighted random walks for graph matching," in *European conference on Computer vision*, 2010.

[32] Z.-Y. Liu and H. Qiao, "GNCCP graduated non convexity and concavity procedure," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 6, pp. 1258–1267, 2014.

[33] Y. Suh, K. Adamczewski, and K. Mu Lee, "Subgraph matching using compactness prior for robust feature correspondence," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[34] O. Duchenne, F. Bach, I.-S. Kweon, and J. Ponce, "A tensor-based algorithm for high-order graph matching," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2383–2395, 2011.

[35] R. Zass and A. Shashua, "Probabilistic graph and hypergraph matching," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[36] Q. Nguyen, A. Gautier, and M. Hein, "A flexible tensor block coordinate ascent scheme for hypergraph matching," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[37] J. Yan, C. Li, Y. Li, and G. Cao, "Adaptive discrete hypergraph matching," *IEEE transactions on cybernetics*, vol. 48, no. 2, pp. 765–779, 2018.

[38] H. J. Chang, T. Fischer, M. Petit, M. Zambelli, and Y. Demiris, "Learning kinematic structure correspondences using multi-order similarities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 1, pp. 1–1, 2017.

[39] S. Rabanser, O. Shchur, and S. Günnemann, "Introduction to tensor decompositions and their applications in machine learning," *Machine Learning*, vol. 98, no. 1-2, pp. 1–5, 2015.

[40] M. Sion *et al.*, "On general minimax theorems." *Pacific Journal of mathematics*, vol. 8, no. 1, pp. 171–176, 1958.

[41] T. H. Haveliwala, "Topic-sensitive pagerank," in *International conference on World Wide Web*, 2002.

[42] F. Han, X. Yang, Y. Deng, M. Rentschler, D. Yang, and H. Zhang, "SRAL: Shared representative appearance learning for long-term visual place recognition," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 1172–1179, 2017.

[43] B. Sheng, C. Shen, G. Lin, J. Li, W. Yang, and C. Sun, "Crowd counting via weighted VLAD on a dense attribute feature map," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 8, pp. 1788–1797, 2016.

[44] S. Zhao, H. Yao, Y. Zhang, Y. Wang, and S. Liu, "View-based 3D object retrieval via multi-modal graph learning," *Signal Processing*, vol. 112, pp. 110–118, 2015.

[45] P. Pinggera, D. Pfeiffer, U. Franke, and R. Mester, "Know your limits: Accuracy of long range stereoscopic object measurements in practice," in *European Conference on Computer Vision*, 2014.

[46] H. Fu, M. Gong, C. Wang, K. Batmanghelich, and D. Tao, "Deep ordinal regression network for monocular depth estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.