Visual Reference of Ambiguous Objects for Augmented Reality-Powered Human-Robot Communication in a Shared Workspace

Peng Gao, Brian Reily, Savannah Paul, and Hao Zhang(✉)

Human-Centered Robotics Lab, Colorado School of Mines,
1500 Illinois Street, Golden, CO 80401, USA
{gaopeng,breily,savannahpaul,hzhang}@mines.edu

Abstract. In shared workspaces, teammates working with a common set of objects must be able to unambiguously reference individual objects in order to effectively collaborate. When teammates are autonomous robots, human teammates must be able to communicate their intended reference object without overtly interfering with their workflow. In human-robot interaction, the problem of visual reference is defined as identifying the specific object referred to by a human (e.g., through a pointing gesture recognized by an augmented reality device), and relating this object to the associated object in the robotic teammate's field of view, thereby identifying the intended object from a set of ambiguous objects. As human and robot teammates typically observe their shared workspace from differing perspectives, achieving visual reference of objects is a challenging yet crucial problem. In this paper, we present a novel approach to visual reference of ambiguous objects that introduces a graph matching-based approach which fuses visual and spatial information of the objects in a shared workspace through augmented reality-powered human-robot communication. Our approach represents the objects in a scene with a graph where edges encoding the spatial relationships among objects and attribute vectors describing each object's appearance associated with each node. Then, we formulate visual object reference for human-robot communication in a shared workspace as an optimization-based graph matching problem, which identifies the correspondence of nodes in graphs built from the human and robot teammates' observations. We conduct extensive experimental evaluation on two introduced datasets, showing that our approach is able to obtain accurate visual references of ambiguous objects and outperforms existing visual reference methods.

Keywords: Visual reference · Human-robot communication · Collaborative perception · Augmented reality · Graph matching

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2020, LNCS 12190, pp. 550–561, 2020.
https://doi.org/10.1007/978-3-030-49695-1_37

1 Introduction

Human-robot teaming and interaction in shared space has been recently attracting significant attention. In such scenarios, autonomous robots work alongside humans in shared workspaces and cooperate with them to complete collaborative tasks together. A critical capability required for human-robot teaming in shared workspaces is effective communication between robots and humans [1]. Recently, augmented reality (AR) has been proposed as a promising solution to improving human-robot communication, offering the revolutionary capability to visualize virtual information with the real world scene through the use of head-mounted displays or hand-held devices [2,3], in order to more intuitively deliver information to a human.

Visual reference of objects in a shared space is essential to enabling AR-powered human-robot communication. Visual reference in human-robot communication is defined as the problem of a robot correctly visually identifying a specific object within the robot's field of view that has been referred to by a human within their own field of view. Figure 1 illustrates a motivating example of visual reference for AR-powered human-robot communication in a shared workspace. A human teammate wears an AR headset and uses a pointing gesture recognized by the AR device to refer to one of the objects in the shared workspace that she wants the robot to pick up. Then, the robot must identify the specific object referred to by the human teammate within its own field of view, before taking an action to manipulate the object. Scenarios that require visual object references are common in many real-world applications, such as

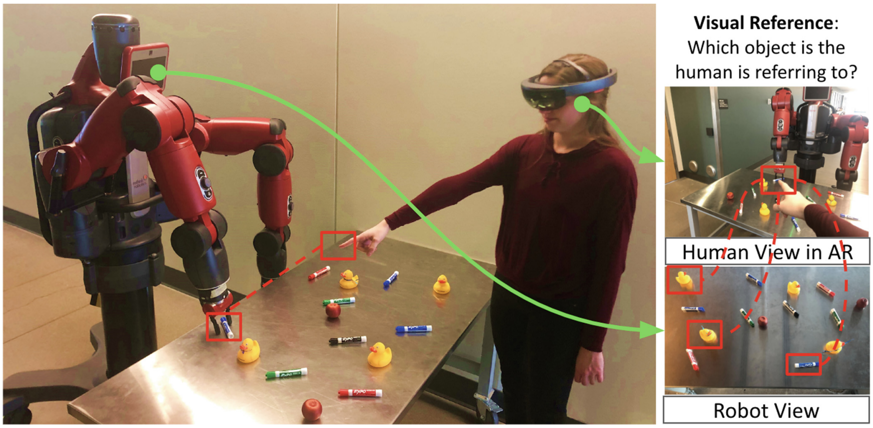


Fig. 1. A motivating example of visual reference of ambiguous objects for AR-powered human-robot communication in a shared workspace. When the human teammate selects an object through a pointing gesture recognized by the AR headset, the robot teammate needs to identify the object referred to by the human within its own field of view, without being confused by other similar objects in the shared space.

assembling, manufacturing, and multi-robot multi-human teaming in field environments, in which many relevant objects exist in shared space.

Due to the importance of visual reference, many methods have been developed to address this problem of identifying a referenced object in multiple views. The first category consists of methods based on visual appearance similarity between objects, e.g., using matched key-points to track the same object in a video [4], utilizing the similarity between region-based global features to identify if two instances of objects are the same [5] or leveraging feature attributes for object re-identification [6]. The second category of existing methods employ a synchronization algorithm to identify the same objects in multiple views by enforcing a circle consistent constraint, e.g., formulating visual reference as a semi-definite problem [7], solving a relaxed convex problem to associate the same object over multiple views [8,9], or formulating this disambiguation as a top-rank eigenvector-decomposition problem [10,11]. The third category of methods to refer to the same object among multiple views consists of approaches based on spatial information among objects, e.g., formulating the visual reference problem as a linear assignment problem which can be solved by the Hungarian [12] or Sinkhorn algorithms [13], identifying a matching node between two graphs by solving the quadratic assignment problem (also known as pairwise graph matching) [14,15], and referring to the same point in multiple point clouds by employing iterative closest point techniques [16].

While these existing methods have obtained promising performance in subareas of visual reference, addressing the problem of disambiguation of object references in a real-world environment in a unified fashion (i.e., incorporating visual and spatial information) is still hard to solve due to two major challenges. The first major challenge for visual object reference is the ambiguity among the multiple objects in shared workspace. Humans and robots typically observe a shared workspace from different points of view. From one teammate's perspective, multiple objects can have a similar appearance (i.e., perceptual aliasing), which will cause the visual appearance and synchronization-based methods to fail, as they require the objects in multiple views to be unique. From different perspectives (e.g., both the robot's and human's), the same object can often exhibit a different visual appearance, such as objects that are different colors on various sides. Thus, using only visual appearance to identify correspondences among objects is insufficient to solve the visual reference problem. The second major challenge for visual object reference is the deformation in the spatial relationships among objects. The position of multiple objects can be obtained from a depth sensor (e.g., using an RGB-D camera). Due to the noise and resolution limit in depth measurement, the position of objects recovered from 2D image coordinates and depth values contain inaccuracies which will lead to deformation in the constructed spatial relationship of objects. Thus, relying solely on spatial relationships among objects is also insufficient to enable effective visual referencing.

In this paper, we propose a graph matching method for visual reference that integrates visual and spatial information describing the objects to identify the correspondence of objects between the robot's view and the human teammate's view. From the robot teammate's perspective, we represent multiple detected objects as a graph, where each node corresponds to a detected object, where edges between nodes describe the spatial distance between objects and an attribute vector associated with each node describes the object's visual appearance. We represent the human teammate's perspective with a similar graph. Thus, our graph representation integrates both visual and spatial information about the detected objects in both observations. Given these two graph representations generated from the robot's and human teammate's observation, we formulate visual reference as a graph matching problem, which uses constrained optimization to identify corresponding objects between two views based on the similarity of the visual and spatial information of the objects encoded in each graph.

The contributions of this research are twofold:

- We formulate visual reference as a graph matching problem, which integrates visual and spatial information of objects into a unified graph representation for correspondence identification between two views to improve representation expressiveness and reduce or remove the ambiguity inherent in visual reference.
- We develop a heuristic optimization algorithm to solve the proposed non-convex graph matching problem, which has no closed-form solution.

The remainder of the paper is organized as follows. In Sect. 2, we introduce the proposed graph matching approach to visual reference of ambiguous objects, including our problem formulation and optimization algorithm. In Sect. 3, we present our experimental results on multiple datasets. Finally, we conclude this work in Sect. 4.

2 The Proposed Approach

Notation. The matrices are denoted as 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 coloumn of $\mathbf{M}$. Vectors are represented as boldface lowercase letters $\mathbf{v} \in \mathbb{R}^n$. In addition, $\mathbf{m} \in \mathbb{R}^{nm}$ is the vectorized form of a matrix $\mathbf{M} \in \mathbb{R}^{n \times m}$, which is a concatenation of each column in $\mathbf{M}$ into a vector.

2.1 Problem Formulation

The problem of visual reference is to identify the same objects in a shared workspace observed by a robot and a human teammate from their respective fields of view. In this work, we design a graph matching-based method to address visual reference.

Given an observation from by a robot or a human teammate, we represent it as a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{F}, \mathcal{S}\}$. In graph $\mathcal{G}$, $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ denotes the node set, which represents the 3D central positions of the objects in the observation, where $\mathbf{v}_i = \{x, y, z\}$ denotes the 3D central position of the i-th object and n is the number of objects. The appearance set $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$ denotes the set of visual appearance features of the objects, where $\mathbf{f}_i$ is a feature vector describing the appearance of the i-th object located at $\mathbf{v}_i$. $\mathcal{S} = \{s_{i,j}, i, j = 1, 2, \ldots, n, i \neq j\}$ denotes the set of spatial relationships among the objects, where $s_{i,j}$ denotes the distance between the i-th object located at $\mathbf{v}_i$ and the j-th object located at $\mathbf{v}_j$.

In visual reference, the objects observed by a robot and a human teammate in their own fields of view can be respectively represented as two graphs $\mathcal{G} = \{\mathcal{V}, \mathcal{F}, \mathcal{S}\}$ and $\mathcal{G}' = \{\mathcal{V}', \mathcal{F}', \mathcal{S}'\}$. Given the graph representations, one way to identify the correspondences of objects in two views is based on visual similarities of objects, which is denoted as $\mathbf{a} = \{\mathbf{a}_{i,i'}\} \in \mathbb{R}^{nn'}$, where $\mathbf{a}_{i,i'}$ represents the similarity between the feature vectors $\mathbf{f}_i \in \mathcal{F}$ of the i the object in graph $\mathcal{G}$ and $\mathbf{f}'_i \in \mathcal{F}'$ of the i' object in graph $\mathcal{G}'$, which can be computed using a dot product of two feature vectors:

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

In order to obtain the correspondence of objects in two different graphs $\mathcal{G}$ and $\mathcal{G}'$ given this appearance similarity, a common way is to solve the following linear assignment problem,

$$\begin{aligned} &\max_{\mathbf{X}} \mathbf{a}^\top \mathbf{x} \\ &\text{s.t.} \quad \mathbf{X} \mathbf{1}_{n' \times 1} \leq \mathbf{1}_{n \times 1}, \mathbf{X}^\top \mathbf{1}_{n \times 1} \leq \mathbf{1}_{n' \times 1} \end{aligned} \tag{2}$$

where $\mathbf{x} = \{\mathbf{x}_{ii'}\} \in \{0,1\}^{nn'}$ denotes the vectorized correspondence matrix $\mathbf{X} \in \{0,1\}^{n \times n'}$, with $\mathbf{X}_{ii'} = 1$ denoting that the i-th node in $\mathcal{V}$ and the i'-th node in $\mathcal{V}'$ are matched, and $\mathbf{1}$ is a vector with all elements equal to 1. In general, the problem in Eq. (2) can be solved by the Hungarian [12] or Sinkhorn algorithms [13]. The constraints in Eq. (2) are designed to enforce the one-to-one correspondence: each row or column in $\mathbf{X}$ can at most have one element equal to 1, and all others are equal to 0.

However, this visual-based matching can not address the ambiguity in object appearance, e.g., two different objects may look similar or the same object observed from different perspectives may look different. Another way to refer the same object in different views is based upon spatial relationships among objects. The distance-based spatial similarity can be denoted as $\mathbf{S} = \{\mathbf{S}_{ii',jj'}\} \in \mathbb{R}^{nn' \times nn'}$, where $\mathbf{S}_{ii',jj'}$ represents the similarity of the distance $s_{i,j} \in \mathcal{S}$ and the distance $s'_{i',j'} \in \mathcal{S}'$, which can be computed by:

$$\mathbf{S}_{ii',jj'} = \exp\left(-\frac{1}{\beta}(s_{i,j} - s'_{i',j'})^2\right) \tag{3}$$

This similarity is computed using the nonlinear exponential function, with β as a hyperparameter, to transfer any non-negative input to an output value between 0 and 1. Then, visual reference can be formulated as the quadratic assignment problem as follows,

Algorithm 1: The proposed algorithm to solve the formulated non-convex optimization problem in Eq. (5).

```
Input  : $\mathbf{S} \in \mathbb{R}^{nn' \times nn'}$ and $\mathbf{a} \in \mathbb{R}^{nn'}$
Output: $\mathbf{X} =\in \{0,1\}^{n \times n'}$
Initialize the correspondence matrix $\mathbf{X}$
Compute $\mathbf{P}$ and $\mathbf{q}$ according to Eq. (6)
while not converge do
    Compute the jump vector $\mathbf{z}$ by Eq. (8)
    Normalize $\mathbf{z}$ using bistochastic normalization
    Update $\mathbf{X}$ with the reweighted jump vector by Eq. (9)
end
Discretize $\mathbf{X}$ using the Hungarian algorithm
return $\mathbf{X}$
```

$$\begin{aligned} &\max_{\mathbf{X}} \mathbf{x}^\top \mathbf{S} \mathbf{x} \\ &\text{s.t.} \quad \mathbf{X} \mathbf{1}_{n' \times 1} \le \mathbf{1}_{n \times 1}, \mathbf{X}^\top \mathbf{1}_{n \times 1} \le \mathbf{1}_{n' \times 1} \end{aligned} \tag{4}$$

Generally, Eq. (4) is NP-hard and can be solved by existing heuristic algorithms [14]. Due to noise in measurement or variations in the sensing environment, the position of objects obtained by sensors is inaccurate, which will lead to deformation in spatial relationships (e.g., distance) among objects. Thus, only using spatial information for visual reference is insufficient.

In this paper, in order to accurately reference the same object in two different views despite the ambiguity in the appearance of objects and the deformation in the position of objects, we formulate visual reference as a graph matching problem defined in an optimization framework, which integrates both the visual and spatial information of objects for visual reference. The final correspondences of the same objects in different views can be obtained by solving the following optimization problem:

$$\begin{aligned} &\max_{\mathbf{X}} \lambda_1 \mathbf{a}^\top \mathbf{x} + \lambda_2 \mathbf{x}^\top \mathbf{S} \mathbf{x} \\ &\text{s.t.} \quad \mathbf{X} \mathbf{1}_{n' \times 1} \le \mathbf{1}_{n \times 1}, \mathbf{X}^\top \mathbf{1}_{n \times 1} \le \mathbf{1}_{n' \times 1} \end{aligned} \tag{5}$$

The first term in Eq. (5) represents the accumulated similarity between visual appearances of the objects in the two graphs, which sums all visual appearance similarities $\mathbf{a}_{i,i'}$ of the feature vectors $\mathbf{f}_i \in \mathcal{F}$ and $\mathbf{f}'_i \in \mathcal{F}'$. The second term denotes the accumulated spatial similarities of the objects in two graphs, which sums all distance similarities $\mathbf{S}_{ii',jj'}$ of two edges $s_{i,j} \in \mathcal{S}$ and $s'_{i',j'} \in \mathcal{S}'$. $\lambda_1 \ge 0$ and $\lambda_2 \ge 0$ are hyperparameters to control the importance of the visual and spatial similarities, which satisfy $\lambda_1 + \lambda_2 = 1$.

2.2 Optimization Algorithm

Since our proposed graph matching formulation is a non-convex optimization problem, which has no closed-form solution, we design an optimization algorithm based on random walks with the reweighted jump technique [14] to solve the proposed optimization problem in Eq. (5). The proposed optimization algorithm is presented in Algorithm 1.

In Step 2, we first convert the similarity matrix in Eq. (5) to a stochastic form $\mathbf{P} \in \mathbb{R}^{nn' \times nn'}$ and $\mathbf{q} \in \mathbb{R}^{nn'}$ as follows:

$$\begin{aligned} \mathbf{P} &= \mathbf{S} / \max_i \sum_j \mathbf{S}_{i,j} \\ \mathbf{q} &= \mathbf{a} / \max_i \mathbf{a}_i \end{aligned} \tag{6}$$

By dividing by the maximum elements, Eq. (6) can normalize the original matrix without losing relative similarity. Then the original formulation in Eq. (5) is converted to the following form:

$$\begin{aligned} &\max_{\mathbf{X}} \lambda_1 \mathbf{q}^\top \mathbf{x} + \lambda_2 \mathbf{x}^\top \mathbf{P} \mathbf{x} \\ &\text{s.t.} \quad \mathbf{X} \mathbf{1}_{n' \times 1} \leq \mathbf{1}_{n \times 1}, \mathbf{X}^\top \mathbf{1}_{n \times 1} \leq \mathbf{1}_{n' \times 1} \end{aligned} \tag{7}$$

Then in Step 4, we define the reweighting jump vector as $\mathbf{z} \in \mathbb{R}^{nn'}$ to jump out of local optima, which is inspired by the PageRank algorithm [14].

$$\mathbf{z}^r = \exp\left(\mathbf{x}^r \circ \mathbf{a} / \max\left(\mathbf{x}^r \circ \mathbf{a}\right)\right) \tag{8}$$

where $\circ$ denotes the entry-wise product and r denotes the r–th iteration. The node appearance similarity $\mathbf{a}$ is introduced to guide the jump toward a direction such that similar objects can be better matched.

In Step 5, we employ bistochastic normalization to normalize each row and column in $\mathbf{z}$ to enforce the one-to-one correspondence.

In Step 6, to facilitate $\mathbf{X}$ to jump out of local optima, $\mathbf{X}$ is updated by:

$$\mathbf{x}^{r+1} = \alpha\left(\mathbf{q} + \mathbf{x}^{r\top} \mathbf{P}\right) + (1 - \alpha)\, \mathbf{z}^r \tag{9}$$

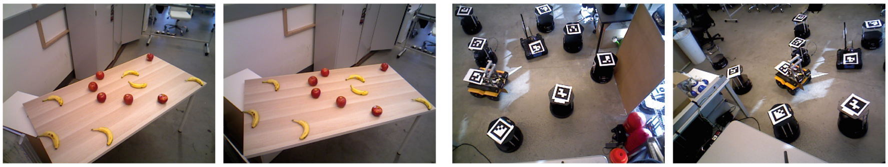
where α is a hyper-parameter that controls the update rate, and $\alpha = 0.3$ in the following experiments.

In Step 8 after convergence, we discretize the matrix $\mathbf{X} \in \mathbb{R}^{n \times n'}$ into binary form by using the Hungarian algorithm. The complexity of the proposed Algorithm 1 is dominated by matrix $\mathbf{S}$, which is $O((nn')^2)$.

3 Experiment

3.1 Experimental Setup

In order to evaluate our proposed visual reference method in AR-powered human-robot communication, we collected two datasets: (1) the multi-view object identification dataset (MOI); and (2) the multi-robot coordination dataset (MRC), seen illustrated in Fig. 2. In each dataset, 30 data instances are used to



(a) MOI (b) MRC

Fig. 2. Illustrations of MOI and MRC datasets.

evaluate our approach. Each data instance contains a pair of RGB-D observations observed by 3D structured-light cameras from two different perspectives. Both of the datasets include the color and HOG features of objects to describe the appearance of objects, 3D position of objects recovered from RGB-D observations and the ground truth of the correspondences of the objects in two perspectives, which are labeled manually in MOI or detected from the QR code in MRC.

We evaluated our method on the MOI and MRC datasets by comparing with three existing methods: (1) **SM** [15], a spatial-based matching method; (2) **RRWM** [14], which uses the distance relationships of objects to identify the correspondence of objects; and (3) **ATT** [6], which uses attribute features to describe object for re-identification. We use each author's original MATLAB code in our comparison. Our visual-spatial graph matching method is also implemented in MATLAB and sets $\lambda_1 = 0.5$ and $\lambda_2 = 0.5$ for evaluation.

We adopt *accuracy*, *precision* and *recall* as metrics to evaluate the performance of visual reference. From the perspective of graph matching [14], given the visual correspondences of the objects, *accuracy* is defined as the number of correct correspondences over the total number of the ground truth of object correspondence. From the perspective of object retrieval [17], *precision* is defined as the fraction of correctly retrieved object correspondences over all retrieved objects, and *recall* is defined as the ratio of retrieved correct correspondences over all of the ground truth of correspondences.

Table 1. Quantitative experimental results on the MOI and MRC datasets.

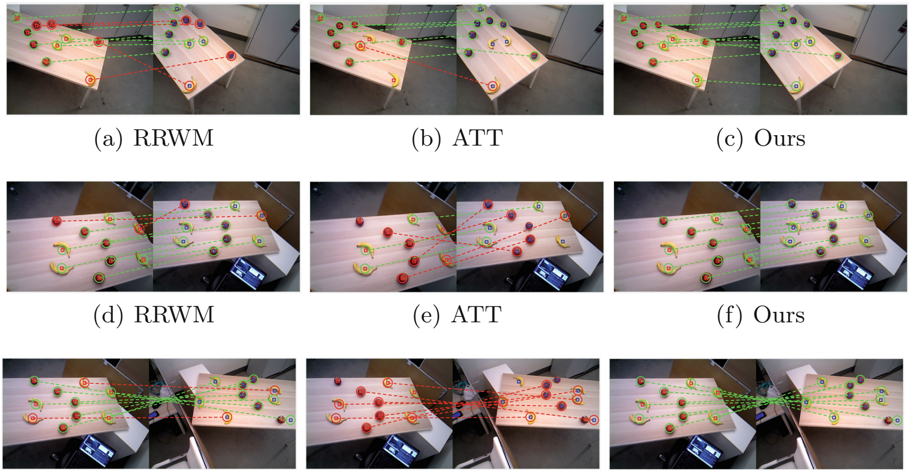
Method	MOI			MRC		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
SM [15]	0.2646	0.2792	0.1874	0.2861	0.2142	0.1926
RRWM [14]	0.2917	0.3440	0.2207	0.3053	0.3262	0.1378
ATT [6]	0.5669	0.6455	0.5579	0.6289	0.6476	0.5211
Ours	**0.8496**	**0.8579**	**0.8056**	**0.8590**	**0.7926**	**0.8434**

3.2 Results on the MOI Dataset

We first evaluate our proposed method on the multi-view object identification dataset. In MOI, the objects are observed from the overhead perspective, in which most classes of the objects have a similar appearance (e.g., multiple apples look similar visually) and the positions of object are inaccurate due to sensing error from the camera angle. Since all the objects are close to each other, the inaccurate positions of object will lead to large deformations in the distance relationships between objects.

The qualitative results are presented in Fig. 3. We can observe that ATT, using only visual information about the objects for visual reference, can only distinguish the objects in different classes that have unique appearances. Due to the existence of ambiguity in objects' appearance, apples in two perspectives are identified incorrectly. The results obtained from the spatial-based method RRWM indicate that using spatial information is robust to the ambiguity in objects' appearance. However, since there exists deformation in the spatial relationships of objects, only using spatial information is still insufficient. Our proposed method integrating both visual and spatial information of objects for visual reference can correctly identify the correspondences of objects in two perspectives and obtain the best performance.

The qualitative results for visual reference in MOI are reported in Table 1. The results indicate that only using spatial-based features (SM and RRWM) causes poor performance for visual reference due to the deformation in the distances between pairs of objects. The visual-based method (ATT) obtains an improved performance but is still worse than our approach due to ambiguity in visual appearance of objects. Our approach achieved the best performance,



(a) RRWM (b) ATT (c) Ours

(d) RRWM (e) ATT (f) Ours

(g) RRWM (h) ATT (i) Ours

Fig. 3. Qualitative experimental results on the MOI dataset with comparison of our proposed method with RRWM and ATT. Green lines denote correct correspondences and red lines denote incorrect correspondences. [Best viewed in color.]

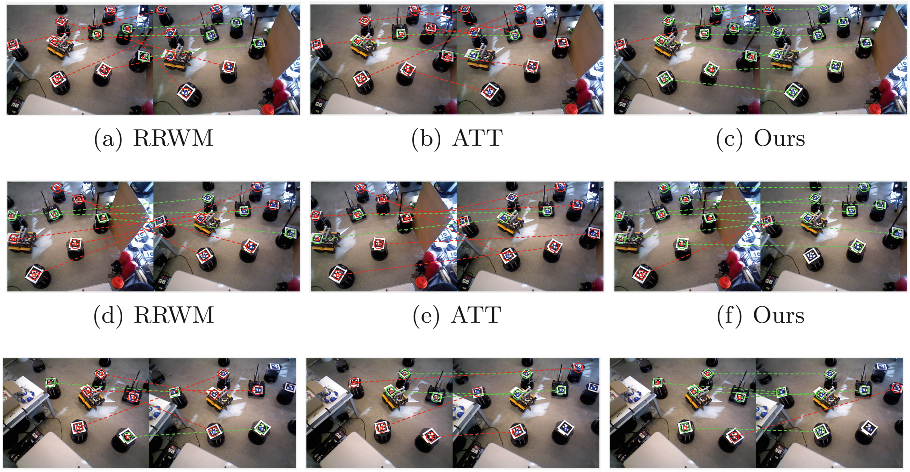
which can address the challenges caused by the ambiguity in visual appearance and deformation in spatial relationships of objects.

3.3 Results on the MRC Dataset

We also evaluate our approach on the multi-robot coordination dataset, which is more challenging compared to the MOI dataset. Most object instances in MRC have an identical appearance, which indicates larger ambiguity in objects' appearance. In addition, the object instances are observed far away from the sensing camera, which leads to larger deformations in objects' spatial relationships. Thus, MRC is a very challenging dataset for the evaluation of visual reference.

To visualize object correspondences, the qualitative experimental results of visual reference for MRC are presented in Fig. 4. Results obtained by the other two best performing methods (RRWM and ATT) are also compared in the figure. We can see that RRWM cannot well identify correspondence of the objects. ATT obtains improved performance, in which unique objects can be correctly identified, but ATT again failed for the objects with identical appearance. As with the MOI dataset, our graph matching approach obtains the best results on object correspondence identification for visual reference.

The quantitative results on MRC dataset are also presented in Table 1. We observe that graph matching methods (SM and RRWM) perform badly, due to the large deformation existed in the spatial relationships of objects. ATT achieves better performance due to the existence of the objects with unique appearance.



(a) RRWM (b) ATT (c) Ours

(d) RRWM (e) ATT (f) Ours

(g) RRWM (h) ATT (i) Ours

Fig. 4. Qualitative experimental results on the MRC dataset with comparison of our proposed method with RRWM and ATT. Green lines denote correct correspondences and red lines denote incorrect correspondences. [Best viewed in color.]

Our approach obtains the best performance on MRC, and outperforms ATT, due to our approach's ability to integrate visual and spatial information about objects.

4 Conclusion

In this paper, we present an approach to enable visual reference of ambiguous objects in a workspace shared between human and robot teammates utilizing augmented reality-powered human-robot communication. We introduce a novel graph matching-based approach that is able to fuse visual and spatial information describing the objects as seen from each teammate's perspective, where these objects as represented with graph edges describing spatial structure and attribute vectors describing visual appearances. We formulate our approach as a constrained optimization problem that identifies the correspondences of nodes in graphs built from both the human and robot's observations. Through extensive evaluation on two datasets, we show that our approach provides accurate visual reference of ambiguous objects from human and robot perspectives, and is able to outperform existing state-of-the-art visual reference methods.

Acknowledgments. This work was partially supported by ARL W911NF-17-2-0181, ARO W911NF-17-1-0447, USAFA FA7000-18-2-0016, DOE DE-FE0031650, IIS-1849348, and NSF CNS-1823245.

References

1. Reardon, C., Zhang, H., Fink, J.: Shaping of shared autonomous solutions with minimal interaction. Front. Neurorobot. **12**, 54 (2018)
2. Williams, T., Szafir, D., Chakraborti, T., Phillips, E.: Virtual, augmented, and mixed reality for human-robot interaction (VAM-HRI). In: ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 671–672 (2019)
3. Reardon, C., Lee, K., Rogers, J.G., Fink, J.: Augmented reality for human-robot teaming in field environments. In: Chen, J.Y.C., Fragomeni, G. (eds.) HCII 2019. LNCS, vol. 11575, pp. 79–92. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21565-1_6
4. Nebehay, G., Pflugfelder, R.: Consensus-based matching and tracking of keypoints for object tracking. In: IEEE Winter Conference on Applications of Computer Vision, pp. 862–869. IEEE (2014)
5. Zhao, R., Oyang, W., Wang, X.: Person re-identification by saliency learning. IEEE Trans. Pattern Anal. Mach. Intell. **39**(2), 356–370 (2016)
6. Zhao, Y., Shen, X., Jin, Z., Lu, H., Hua, X.-S.: Attribute-driven feature disentangling and temporal aggregation for video person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4913–4922 (2019)
7. Leonardos, S., Zhou, X., Daniilidis, K.: Distributed consistent data association via permutation synchronization. In: IEEE International Conference on Robotics and Automation (2017)

8. Hu, N., Huang, Q., Thibert, B., Guibas, L.J.: Distributable consistent multi-object matching. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
9. Zhou, X., Zhu, M., Daniilidis, K.: Multi-image matching via fast alternating minimization. In: IEEE International Conference on Computer Vision (2015)
10. Maset, E., Arrigoni, F., Fusiello, A.: Practical and efficient multi-view matching. In: IEEE International Conference on Computer Vision (2017)
11. Pachauri, D., Kondor, R., Singh, V.: Solving the multi-way matching problem by permutation synchronization. In: Advances in Neural Information Processing Systems (2013)
12. Almohamad, H., Duffuaa, S.O.: A linear programming approach for the weighted graph matching problem. IEEE Trans. Pattern Anal. Mach. Intell. **15**(5), 522–525 (1993)
13. Wang, R., Yan, J., Yang, X.: Learning combinatorial embedding networks for deep graph matching. In: IEEE International Conference on Computer Vision (2019)
14. Cho, M., Lee, J., Lee, K.M.: Reweighted random walks for graph matching. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6315, pp. 492–505. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15555-0_36
15. Leordeanu, M., Hebert, M.: A spectral technique for correspondence problems using pairwise constraints. In: Tenth IEEE International Conference on Computer Vision, ICCV 2005, vol. 2, pp. 1482–1489. IEEE (2005)
16. Sobreira, H., et al.: Map-matching algorithms for robot self-localization: a comparison between perfect match, iterative closest point and normal distributions transform. J. Intell. Robot. Syst. **93**(3–4), 533–546 (2019)
17. Suh, Y., Adamczewski, K., Mu Lee, K.: Subgraph matching using compactness prior for robust feature correspondence. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5070–5078 (2015)