World State-dependent Action Graph: A Representation of Action Possibility and Its Variations in Real Space based on World State

Yosuke Kawasaki[1] [a] and Masaki Takahashi[2] [b]

[1] *Graduate School of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan*

[2] *Department of System Design Engineering, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan*

Keywords: Service Robot, Environmental Representation, Action Possibility.

Abstract: For intelligent systems, it is important to understand the action possibility for agent in real space. As the action possibility varies with the subsystem configuration of the agent and its states, the possibilities should be understood based on the world state comprising the agent's state as well as the environmental state. However, most conventional methods consider only the environmental state. Therefore, this study proposes a world state-dependent action graph based on knowledge representation using scene graphs which allows the capturing of the action possibility of agents, which implies the feasible actions and their positions in real space, and their recursive variations depending on the world state. Moreover, the effectiveness of the proposed method was verified with simulations, assuming a coffee shop environment.

1 INTRODUCTION

The feasible actions for intelligent systems such as robots and other agents such as humans or robots in real space are required to be understood. In particular, understanding the feasible actions facilitates various kinds of reasoning, such as planning, understanding instructions from humans, and behavior prediction.

The feasible actions for the agent depend on the environment state such as the objects and scenes in the environment. Also, the feasible actions for the agent depend on the subsystem configuration of the agent and their states. For instance, if an agent's gripper is grasping an object, the gripper cannot feasibly grasp onto other objects. In summary, the feasible actions for the agent depend on the world state, i.e., the states of both agent and environment. On the contrary, the world state varies with the actions executed by the agent. Thus, the action of the agent alters the world state, which consequently modify the agent's feasible actions. To understand the feasible action including its variations, the mutual interaction between the feasible action according to the world state and the action executions involving the varying world state.

Conventionally, several methods of affordance estimation have been proposed for associating actions to objects and scenes in an environment, such as the image segmentation method based on affordance(Do et al., 2018; Chu et al., 2019). Then, the mutual interaction between the feasible actions and the action executions are represented by using conventional symbolic representations such as STRIPS and PDDL(Fikes and Nilsson, 1971; Fox and Long, 2003). However, existing studies do not focus on understanding the feasible actions in real space, and it can be actually difficult for the agent to execute associated actions.

As an approach for understanding the action possibility, which implies the feasible actions and their positions in real space, certain methods have been proposed to map the feasible actions into real space. One such method constitutes the action graph(Kawasaki et al., 2021), which is an environmental representation method pivoted on the feasible actions in real space. In this graph, the edges indicate actions and the nodes denote the executable locations of the actions to represent the connections between actions. Specifically, the multilayered graph structure of the action graph enables the representation of variations in action possibility. However, the action graph handles not the agent's state but the configuration of the subsystem. In addition, only once change of the action possibility is taken into account, making it difficult to represent the mutual interaction between the action possibility and the action.

[a] https://orcid.org/0000-0002-3076-3258

[b] https://orcid.org/0000-0001-8138-041X

Kawasaki, Y. and Takahashi, M.
World State-dependent Action Graph: A Representation of Action Possibility and Its Variations in Real Space based on World State.
DOI: 10.5220/0011298000003271
In *Proceedings of the 19th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2022)*, pages 459-466
ISBN: 978-989-758-585-2; ISSN: 2184-2809

By extending the action graph, this study aims to understand the action possibility and its variations in real space based on the world state. The requirements for constructing an action graph considering the world state are stated as follows: 1) Construction of a framework for symbolic reasoning to understand the feasible actions and its effects on the world state. 2) Representation of mutual interaction between the action possibility and action execution involving the variations in the world state.

Therefore, to fulfill the objective, we propose a world state-depended action graph (WDAG) based on knowledge representation with scene graphs. In particular, the scene graph is an environmental representation that is compatible with the world state and contains both geometric and semantic information. For this reason, we adopted the scene graph to represent knowledge about action, specifically the preconditions for actions and the changes in the world state caused by the action. Additionally, the WDAG represents the mutual interaction between the action possibility and the action in a recursive multilayered graph structure.

2 RELATED WORK

Our work is related to the study of understanding feasible actions.

As an approach to understanding feasible actions, segmentation of images based on object affordance has been proposed (Do et al., 2018; Chu et al., 2019). The AffordanceNet is a deep learning approach that assigns affordance to each pixel of the object in real-time (Do et al., 2018). In addition, T. Lüddecke et al. proposed a method to rate the action plausibilities in accordance to the object set present in the image (Lüddecke and Wörgötterr, 2020). In the research field of predicting human activity, the consideration of the agent's state can improve the performance of the prediction (Koppula and Saxena, 2016; Liang et al., 2019). Thus, ATCRF is proposed to obtain a distribution for predicting human activities based on their posture and the surrounding environment (Koppula and Saxena, 2016). Nonetheless, these studies do not consider the possibility of action in real space.

Thus, methods have been proposed to map the action possibility onto real space for understanding the action possibility in real space(Ziaeetabar et al., 2017; Rhinehart and Kitani, 2016). In particular, the multilayer environmental affordance map is proposed to represent the traversability into the grid map following the arrangement and affordance of the object set in the environment (Ziaeetabar et al., 2017). Moreover, an action map is proposed, wherein the actions other than the movement are mapped in real space (Rhinehart and Kitani, 2016). However, these methods do not consider the variations in action possibility occurring as a consequence of performing the action. The most relevant study pertains to the action graph, representing the action possibility and its variations in real space (Kawasaki et al., 2021). We propose an action graph based on the agent's state, which is not typically considered, to capture the feasible actions depending on the world state in the real world.

3 WORLD STATE-DEPENDENT ACTION GRAPH

3.1 Overview

WDAG G^a is an environmental representation that depicts an agent's action possibility in a world state W along with its variations.

The world state $W = \{o | o \in R \,\mathrm{or}\, o \in O\}$ can be described as a set of N^o objects comprising the robot state R and environment state O. The object o details the following information:

- l^o: Label of the object
- p^o: Position of the object in a two-dimensional absolute coordinate system
- θ^o: Direction of the object in a two-dimensional absolute coordinate system
- s^o: Shape data of the object comprising geometric primitives such as square or circle and size.

The objects in the robot state $R = \left\{o | p^o = p^R\right\}$ are whose position p^o is p^R, e.g., a manipulator, speaker, or an object grasped by the manipulator, where p^R denotes a parameter reflecting the robot's position for convenience. Conversely, the elements of the environmental state $E = \left\{o | p^o \neq p^R\right\}$ include the objects o with their position p^o indicated as not p^R but as a three-dimensional coordinate in real space.

More importantly, the WDAG G^a defines a pair of sets $G^a = (N^a, E^a)$, where N^a and E^a denote the set of nodes and edges. The nodes, n^a, represent the positions in which the robot can perform actions in real space. Furthermore, the actions a are assigned to the edges e, and a transition between the nodes implies the performance of the set action a.

The WDAG G^a comprises multiple action possibility graphs $G^a_{1 \sim N^W}$. In particular, the action possibility graph G^a_i describes feasible actions for the agent as well as the positions corresponding to the performance of the actions based on the world state W_i. In

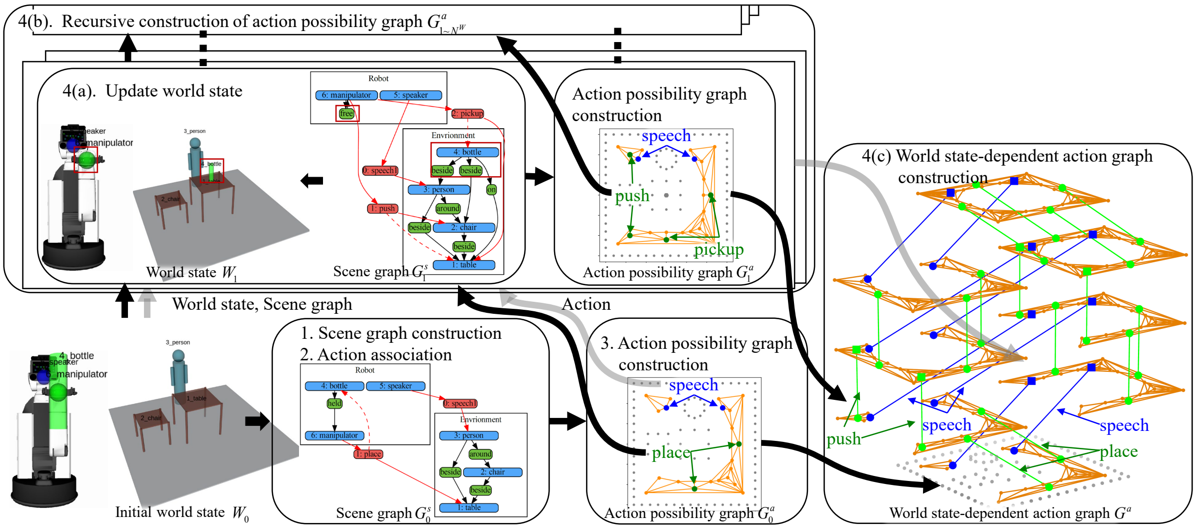


Figure 1: Pipeline diagram of WDAG construction.

the action possibility graph, the edges represent actions excluding the transition of the world state, such as the movement action in real space. On the other hand, the feasible action involving the state transition of the world is denoted by the edges connecting a prior action possibility graph G^a_i with a posterior one G^a_j to the state transition.

Thus, a symbolic inference associating the actions to the world state is required for constructing the action graph. Therefore, we construct a scene graph that includes both the symbolic and geometric information from the world state and incorporates them into the inference. The scene graph G^s defines a pair of sets $G^s = (N^s, E^s)$, where N^s and E^s denote the set of nodes and edges, respectively. The objects o in the world state W correspond to the nodes in the G^s. Furthermore, the edges reflect the spatial relation between the objects.

The pipeline diagram of the automatic construction of the WDAG is illustrated in Fig. 1.

1. Construction of the scene graph G^s based on the world state W to derive the geometric and symbolic information of the W(Section 3.3)
2. Association of actions with the scene graph G^s based on the prior action knowledge (Sections 3.2 and 3.4)
3. Construction of action possibility graph G^a_0 based on the scene graph G^s and associated actions (Section 3.5)
4. Recursive construction of action possibility graphs (Section 3.6)
   (a) Updating scene graph and world state according to the effect of feasible action
   (b) Recursive construction of action possibility graphs based on the updated world state
   (c) Construction of WDGE G^a by connecting the action possibility graphs

Steps 1, 2, 4(a), and 4(b) were modified or added from the conventional method of constructing action graphs.

3.2 Action Knowledge Representation

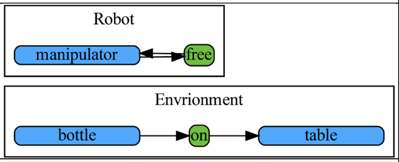
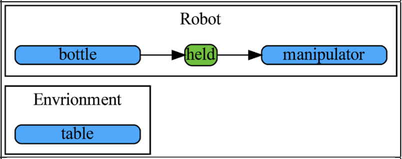
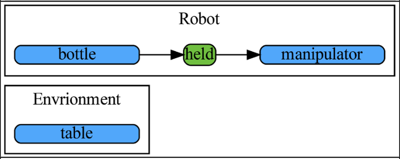
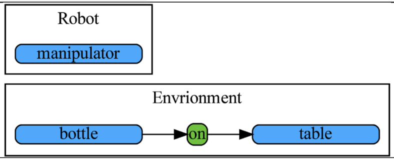
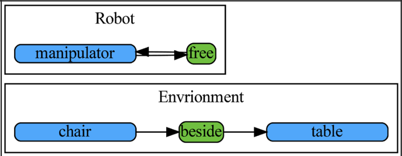
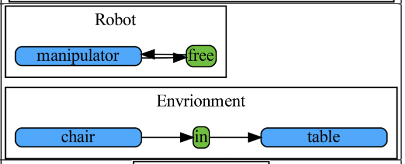
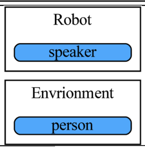
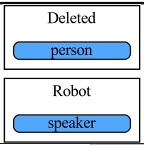
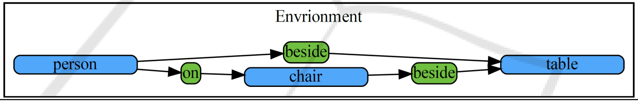
The prior knowledge of the actions is represented using the scene graphs that are compatible with the world state and symbolic reasoning.

Each property of the action is described below.

- l^a: Label of the action
- S^a_p: Scene graph as affordance that affords this action
- d^a: Appropriate distance from the target object to execute the action
- S^a_e: Scene graph after execution of action

The examples of action labels l^a include "movement," "pick up" (to pick up the bottle), "place" (to place the bottle on the table), "push"(to push the chair underneath the table), "speech"(to ask a person to allow passing movement), and "do not disturb" (to dismiss all actions). In particular, the "do not disturb" action is associated with situations involving a seated customer for courteous service. The prior knowledge about action are detailed in Table 1. In this study, we assume that robots can acquire the prior knowledge about action with the affordance classification and the action understanding (Chu et al., 2019; Dreher et al., 2019).

Table 1: Prior knowledge of action.

label	precondition	appropriate distance	effect
pick up	 Robot manipulator free Envrionment bottle on table	0.5m	 Robot bottle held manipulator Envrionment table
place	 Robot bottle held manipulator Envrionment table	0.5m	 Robot manipulator Envrionment bottle on table
push	 Robot manipulator free Envrionment chair beside table	0.7m	 Robot manipulator free Envrionment chair in table
speech		0.4m	
don’t disturb	 Envrionment person on beside chair beside table		

3.3 Scene Graph Construction

A scene graph is constructed based on the world state obtained beforehand by semantic slam, etc. The scene graph G^s is a directed graph in which the nodes indicate objects and edges denote relations based on the world state W. The adjacency matrix of G^s is derived by classifying the relational position into the relation label r for every combination of two objects in the world state. Generally, a positional relation r labels a relative position from a larger object to a smaller one.

The considered labels of positional relations are “around” (distance A from B is 1.0– 2.0 m), “beside” (distance A from B is 0–1 m), “in,” “on,” “held,” and “free” based on previous research (Kim et al., 2019; Armeni et al., 2019). In particular, “held” and “free” are positional relations related to the manipulator. In the case the positions of a manipulator and the other object are the same, the relative position is labeled “held.” A self-looping relation labeled as “free” is associated with the manipulator, if the manipulator does not relate to any object.

3.4 Association of Actions with Scene Graph

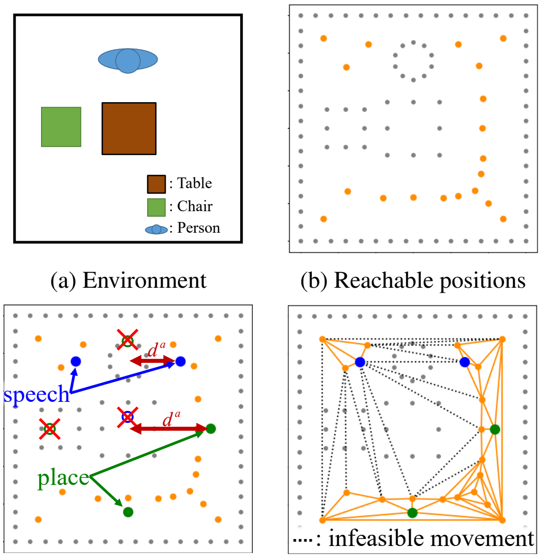
The feasible actions for the agent depend on the world state W comprising the agent state R and environmental state E. Therefore, the feasible actions are associated with the world state based on the precondition of each action in the prior knowledge. Accordingly, the subgraph isomorphism problem is applied to the action association. The subgraphs of the world-state scene graph G^s are isomorphic to the precondition scene graph of an action, which were associated with the action and determined using the VF2 algorithm (Cordella et al., 2004). Thus, multiple actions can be associated with the same object.

3.5 Construction of Action Possibility Graph

3.5.1 Sparse Representation of Reachable Positions

The positions sparsely representing the accessible locations of the agent are determined as nodes focused on the movement.

First, the object footprints including the wall are expressed using a set of points in a two-dimensional



(a) Environment (b) Reachable positions

(c) Action positions (d) Connecting nodes

Figure 2: Construction of action possibility graph.

space. Thereafter, the generalized Voronoi diagram is generated by setting the base points of the diagram equal to the object points. As the generalized Voronoi diagram contains untraversable edges and nodes inaccessible by the robot, the invalid nodes and edges were deleted with reference to the robot's footprint. Furthermore, neighboring nodes are merged together to more efficiently represent the space, as portrayed in Fig. 2b.

3.5.2 Adding Appreciate Action Positions for Associated Actions

The nodes are added to positions considered appropriate for executing the actions associated with the world state, as depicted in Fig. 2c. In particular, certain requirements that ensure the suitability of a given position to perform the action include an appropriate distance to the target object o_t^a and accessibility to the target object position o_t^a based on the agent's footprint and surrounding obstacles. The position candidates suitable for executing the actions are determined based on the appropriate distance d_t^a to the target object o_t^a in the prior knowledge of the action, including the position and posture of the target object o_t^a. In this study, the candidates are simply determined as the four points on the circle whose center denotes the position of the target object and radius is d^a, as depicted in the figure. Thereafter, among the candidates, the nodes are selected if it did not overlap with any object in the environment, and if there is no object situated between itself and the target object o_t^a.

3.5.3 Connecting Nodes

The edges connecting the nodes are determined based on inspecting the movement feasibility through geometric reasoning.

Subsequently, the edge candidates are obtained by applying Delaunay triangulation to the determined nodes. In addition, the edges are adopted by reviewing the intersection based on the object placement in the environment and the agent's footprint. Ultimately, an action possibility graph is obtained, as illustrated in Fig. 2d.

3.6 Recursive Construction of Action Possibility Graph

The actions involving the state transitions of the environment are represented with directed edges that connect the action possibility graphs before and after the state transition. Thus, the exhaustive understanding of the connections between the action possibility graphs is essential to represent the mutual interaction between the action possibility and actions in WDAG. Accordingly, the construction of the action possibility graph in the world state and updating the world state by executable actions are recursively repeated to derive the exhaustive action possibility graphs. Thereafter, the WDAG is obtained by connecting each action possibility graph with edges indicating the action, which is the cause of variation in the action possibility.

The specific process of constructing the WDAG is described below.

3.6.1 Update Scene Graph and World State According to Effect of Action

Herein, the action possibility graph derived in Section 3.5 is used as a starting point. First, the prior scene graph is updated based on each action associated with the obtained action possibility graph, and the posterior scene graph to the action is obtained, as depicted in Fig. 3. The subgraph associated with the action in the prior scene graph is adjusted according to the influence of the action in the prior knowledge. Subsequently, the new scene graphs are obtained for the number of associated actions.

Thereafter, the nodes of the scene graph, i.e., the world states, are updated based on the freshly obtained scene graph, as portrayed in Fig. 3. In particular, the object positions are updated following the relationship altered by the action. Overall, the updated positions of the object are determined following similar rules as those applied for constructing the scene

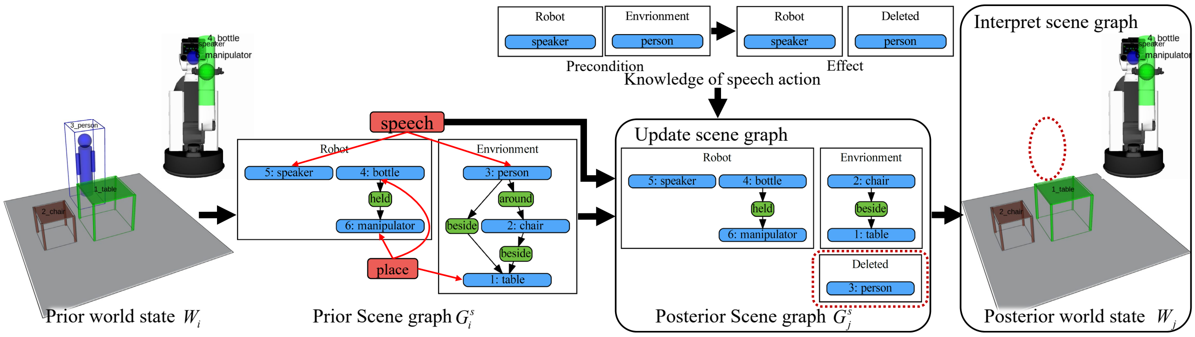


Figure 3: Updating scene graph and world state based on action knowledge.

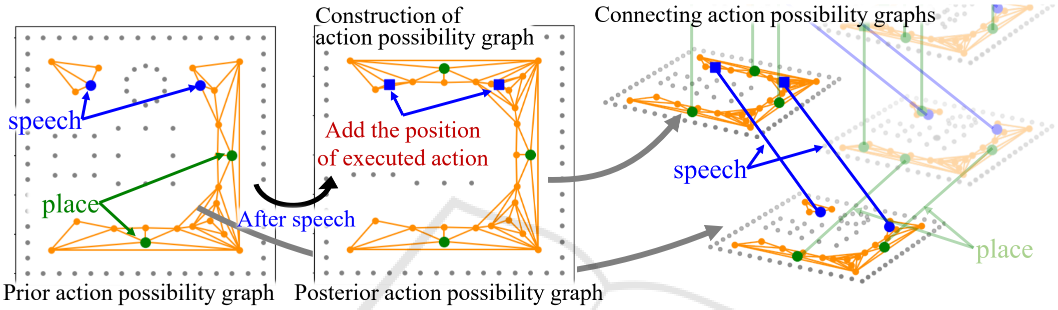


Figure 4: Connecting action possibility graphs.

graph based on the relative position classification algorithm.

3.6.2 Construction of Action Possibility Graph

The subsequent step involves the construction of the new action possibility graphs based on the updated scene graph and world state. Predominantly, the action possibility graph is constructed based on the process described in Sections 3.4 and 3.5. The method was modified with the addition of a node to be considered. Moreover, the executive positions of the actions altering the prior action possibility graph are added as nodes in the new action possibility graph, which enabled the connection of edges before and after the adjustment of the graphs, as shown in Fig. 4.

3.6.3 Iterative Processing

The above process is repeated for each new action possibility graph and its associated actions. In this iteration, as the same scene graph may occur in various backgrounds, they are considered as a single scene graph. This integration limits the number of world states to be considered. In particular, the iteration is terminated when the actions associated with all the graphs were processed. Eventually, all possible versions of the action possibility graphs are obtained.

3.6.4 Connecting the Action Possibility Graphs

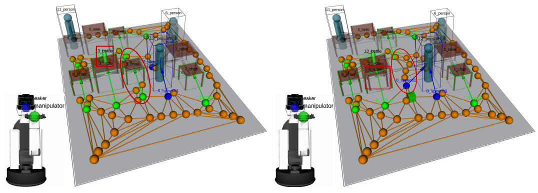
The multilayered graph is constructed by connecting the action possibility graphs with edges. Specifically, the action possibility graphs before and after the adjustments are connected using the edges assigned to the actions causing the modification, as exemplified in Fig. 4. The edges are directed from the prior graph to the posterior graph. Ultimately, the obtained multilayered graph is WDAG.

4 EXPERIMENT

To verify the effectiveness of the proposed method, we qualitatively evaluated the following terms: 1) The proposed method can capture the action possibilities depending on the agent's state. 2) The proposed method can capture the variations in the action possibilities caused by the agent's actions.

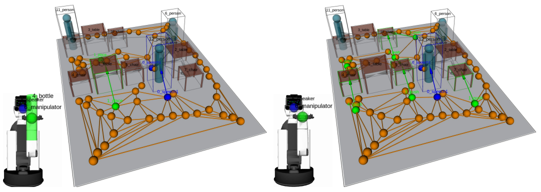
4.1 Setting

In this experiment, we assumed that an agent is HSR (Human support robot), a mobile manipulator, and the environment was similar to that of a café. As displayed in Fig. 6, the target world state included two world states with distinct agent states. In world state



(c) G_2^a (d) G_6^a

Figure 5: Mutual interaction between action possibility and action.



(a) World state 1 (G_0^a) (b) World state 2

Figure 6: Difference of action possibility graph depending on agent's state.

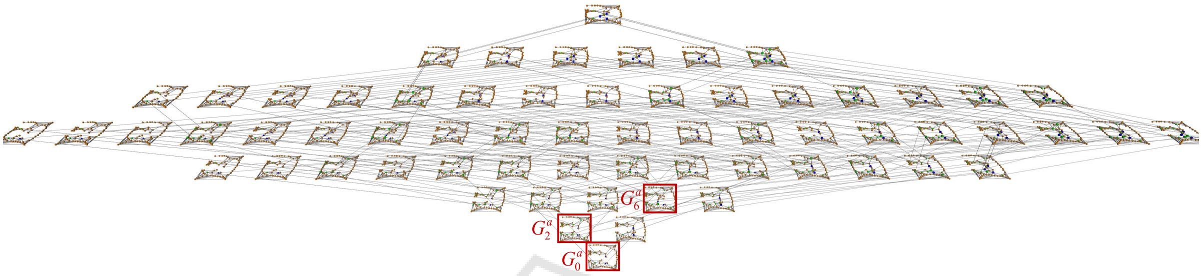


Figure 7: World state-dependent action graph based on world state 1.

1, the agent grasped the bottle, and in world state 2, the agent grasped nothing.

4.2 Result

First, the action possibility graphs constructed based on the world states with the two distinct agent's states are displayed in Fig. 6. The action possibility graphs represent feasible actions for the agent and their positions in real space. According to Fig. 5a, the feasible actions involved placing the bottle on the table and speech toward the person in the world state 1. In comparison, pushing the chair and speech to the person are feasible in the world state 2 according to Fig. 5b. Thus, the proposed method can capture the action possibilities of the agent based on the state of its manipulator.

The WDAG constructed based on the world state 1 is illustrated in Fig. 7. As the WDAG comprised 64 action possibility graphs, the robot could realize 64 world states. In WDAG, the variations in the action possibility graph in case the agent executes the "place" and "push" from the initial state $\left(G_0^a \Rightarrow G_2^a \Rightarrow G_6^a\right)$ are displayed in Fig. 6 The comparison of G_0^a and G_2^a displayed in Fig. 5c revealed that "placing" the bottle on the table enabled the agent to "push" the chair in a new manner. Moreover, the comparison of G_0^a and G_2^a unveiled that the agent could render the location of the chair traversable by pushing the chair, as displayed in Figs. 5c and 5d. Furthermore, the connection of the action possibility graphs displayed that the variations of the world state and the action possibility caused by actions such as that a blocked passage by a chair can be rendered traversable. In summary, the proposed method could capture the mutual interaction between the action and its possibility in the recursive multilayered structure.

Individual action possibility graphs in WDAG represent feasible actions and their positions in real space based on geometric reasoning. Also, the connection of the action possibility graph in WDAG represents the possible world states by executing actions, as well as the conventional symbolic representations. Thus, it is expected that task planning based on WDAG will derive more suitable sequences of actions according to the evaluation index because WDAG combines both symbolic and geometric inference and can comprehensively capture feasible action sequences in real space.

5 CONCLUSION

The understanding of feasible actions for agents requires further detail to realize intelligent systems in real space. Essentially, the feasible actions for the agent vary with the subsystem configuration of the agent and their states. Therefore, the feasible actions depending on the world state that comprises the agent's state and environmental state should be comprehended with clarity. Conventionally, the mutual interaction between the feasible actions and the action executions are represented by using symbolic representations. However, the approaches did not consider the feasibility of actions.

This study aimed to understand the action possibility, which implies the feasible actions and their positions in real space, and its variations in real space based on the world state. To archive the objective, we proposed a WDAG based on knowledge representation using scene graphs. In particular, we adopted the scene graph to represent the knowledge of action, because it is an environmental representation that is compatible with the world state and contains both geometric and semantic information. In addition, the WDAG represented the mutual interaction between the action and its possibility in a recursive multilayered graph structure. Accordingly, a construction method of an action graph was established based on the scene graph-based representation of action effects and a recursive multilayered graph structure. This allowed the capturing of the action possibility of agents and the recursive variations of the action possibility depending on the world state. The effectiveness of the proposed method was verified by simulation, assuming a coffee shop environment. Moreover, the following two points were verified. 1) WDAG represents the action possibility in real space based on the world state. 2) WDAG represents the variations in the action possibility caused by the agent's action on the recursive multilayered structure.

In future, we will validate the effectiveness of WDAG in practice by implementing a planning method of action sequences based on WDAG and applying to task plannings in real space. Task planning based on WDAG is expected to yield more efficient plans, such as plans with shorter movement distances, by considering geometric information such as object placement.

ACKNOWLEDGMENTS

This study was supported by the Core Research for Evolutional Science and Technology (CREST) of the Japan Science and Technology Agency (JST) under Grant Number JPMJCR19A1.

REFERENCES

Armeni, I., He, Z.-Y., Zamir, A., Gwak, J., Malik, J., Fischer, M., and Savarese, S. (2019). 3d scene graph: A structure for unified semantics, 3d space, and camera. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5663–5672.

Chu, F.-J., Xu, R., Seguin, L., and Vela, P. A. (2019). Toward affordance detection and ranking on novel objects for real-world robotic manipulation. *IEEE Robotics and Automation Letters*, 4(4):4070–4077.

Cordella, L., Foggia, P., Sansone, C., and Vento, M. (2004). A (sub)graph isomorphism algorithm for matching large graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1367–1372.

Do, T.-T., Nguyen, A., and Reid, I. (2018). Affordancenet: An end-to-end deep learning approach for object affordance detection. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5882–5889.

Dreher, C. R., Waechter, M., and Asfour, T. (2019). Learning Object-Action Relations from Bimanual Human Demonstration Using Graph Networks. *IEEE Robotics and Automation Letters*.

Fikes, R. E. and Nilsson, N. J. (1971). Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2(3):189–208.

Fox, M. and Long, D. (2003). PDDL 2.1: An extension to pddl for expressing temporal planning domains. *Journal of artificial intelligence research*, 20:61–124.

Kawasaki, Y., Mochizuki, S., and Takahashi, M. (2021). Astron: Action-based spatio-temporal robot navigation. *IEEE Access*, 9:141709–141724.

Kim, U.-H., Park, J.-M., Song, T.-J., and Kim, J.-H. (2019). 3-D Scene Graph: A Sparse and Semantic Representation of Physical Environments for Intelligent Agents. *IEEE Transactions on Cybernetics*, pages 1–13.

Koppula, H. S. and Saxena, A. (2016). Anticipating human activities using object affordances for reactive robotic response. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):14–29.

Liang, J., Jiang, L., Niebles, J. C., Hauptmann, A. G., and Fei-Fei, L. (2019). Peeking into the future: Predicting future person activities and locations in videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5725–5734.

Lüddecke, T. and Wörgötterr, F. (2020). Fine-grained action plausibility rating. *Robotics and Autonomous Systems*, 129:103511.

Rhinehart, N. and Kitani, K. M. (2016). Learning action maps of large environments via first-person vision. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 580–588.

Ziaeetabar, F., Aksoy, E. E., Wörgötter, F., and Tamosiunaite, M. (2017). Semantic analysis of manipulation actions using spatial relations. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4612–4619.