Appendix 3: Ambiguous Instruction Task Planning

Task Description

I will provide you with a picture of a scene, an analysis that includes object attribute analysis, object operability analysis, scene and human intent analysis, and an ambiguous instruction generated from this analysis. You need to decompose the ambiguous instruction into a sequence of atomic instructions that can be executed by a two-fingered gripper robot.

The definition of an atomic instruction is as follows:

An Atomic Instruction is the most basic, indivisible operational command. It is highly specific and executable, requiring no additional reasoning or decision-making. It must contain the following three elements:

(1) A clearly specified object, such as the red cup on the table.

(2) A concrete action; usable actions include grasp, place, push, pull, drag, rotate, and press.

(3) A precise target state or location, such as the oven door being open.

The action library for the two-fingered gripper robot is defined as follows:

Action 1: Grasp

- **Parameter:** Object A
- **Description:** Securely grip the specified object with the gripper.
- **Preconditions:** The gripper is empty, Object A is graspable, the gripper has moved within range to make contact with Object A.
- **Effects:** Object A is grasped, the gripper is not empty.

Action 2: Place

- **Parameters:** Object A, Location A
- **Description:** Place the held object at the specified location.
- **Preconditions:** Object A is grasped, the gripper has reached Location A.
- **Effects:** Object A is not grasped, Object A is at Location A, the gripper is empty.

Action 3: Push/Pull

- **Parameters:** Object A, Start Location, Target Location
- **Description:** Push or pull an object to move it from the start location to the target location.
- **Preconditions:** The gripper has moved within range to make contact with Object A, Object A is at the Start Location.
- **Effects:** Object A is at the Target Location.

Action 4: Rotate (Button)

- **Parameters:** Object A, Start Position, End Position
- **Description:** Rotate an object (like a knob) from a start position to an end position, such as rotating from ON to OFF.
- **Preconditions:** The gripper has moved within range to make contact with Object A, Object A is at the Start Position.
- **Effects:** Object A is at the End Position.

Action 5: Press

- **Parameter:** Object A
- **Description:** Move the gripper to Object A and apply appropriate downward or forward force to complete a pressing action on Object A.
- **Preconditions:** The gripper has moved within range to make contact with Object A, the gripper is empty.
- **Effects:** Object A is pressed.

(Some action skills are omitted here.)

The image, related analysis, and instruction are as follows: {image, text}

Chain of Thought

1. Analyze the target scene:

   Based on the image, the related analysis, and the ambiguous instruction, analyze the specific intent of the instruction. Output a clear and specific description of the final scene to be achieved (including how objects should be placed and arranged).

2. Identify the objects to be manipulated:

Based on the final scene, infer which objects need to be operated on and list them in this step.

3. Generate atomic instructions:

   Based on the intent of the ambiguous instruction, the final scene, and the objects to be manipulated, generate a step-by-step sequence of clear atomic instructions. Each instruction must include a specific object, an action selected from the action library, and a precise target state or location. Output the generated sequence of atomic instructions.

4. Verify the generated atomic instruction sequence:

   Combining the image and general knowledge, check the generated sequence of atomic instructions step-by-step. For each step, state the following:

   - ① State the current scene.
   - ② State the reason for choosing this action.
   - ③ Verify that all [Preconditions] for the chosen action are met.
   - ④ Describe the scene after execution based on the action's [Effects].

   Repeat this process until the target state is achieved to verify the correctness of the generated atomic instruction sequence. Output the analysis for each step. If verification fails, return to step 3 to regenerate the sequence.

5. **Output the complete generated atomic instruction sequence.**