Legend: OpenVLA | VLA ft. Text Trace | VLA ft. Visual Trace

Success Rate (%)
- VLA ft. Visual Trace: 47.7%
- VLA ft. Text Trace: 42.6%
- OpenVLA: 40.2%

You are given an image of robot's observation. Additionally, you are given the movement information of 5 points in the image. Over the last 6 frames, the 5 points moved as follows:

Point 1 moved through positions: [244,80], [200,115], [177,31], ...
Point 2 moved through positions: [218,96], [200,122], [185,10], ...
Point 3 moved through positions: [237,118], [208,42], [161,79], ...
Point 4 moved through positions: [225,49], [200,127], [171,72], ...
Point 5 moved through positions: [251,93], [200,103], [181,91], ...

Now what the robot should do to pick up the coke can?