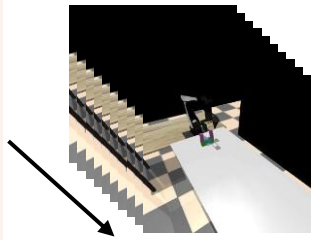
Data:

10 videos

.....



<Prompt>

You are the "Embodied Control Task Decomposition Expert". You will perform cross-video alignment for 10 teaching videos of the same task and automatically determine the reasonable number of stages……

Response:

Gemini

<Analyze>

First, I need to comprehensively analyze the developer role and user task requirements provided by the user…

<Generation>

```
{
"task_name": "block_bridge_place",
"task_description": "Build bridge: place two bars,
  fill with blocks",
"num_stages": 74,
"triplets": {
  "stage_0": {
    "positive": "Gripper above bridge structure;
      fingers open; ready to operate",
    "negative": "Gripper far from all structures;
      fingers closed; static",
    "hard_negative": "Gripper beside bridge structure;
      fingers open; ready"
  },
                          ......
```

Role:

Developer

User

Stage 0 | Stage 1 | ..... | Stage N

Segment 0 | Segment 1 | Segment N