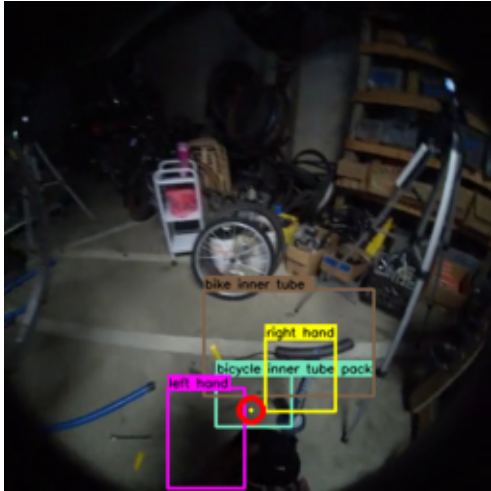


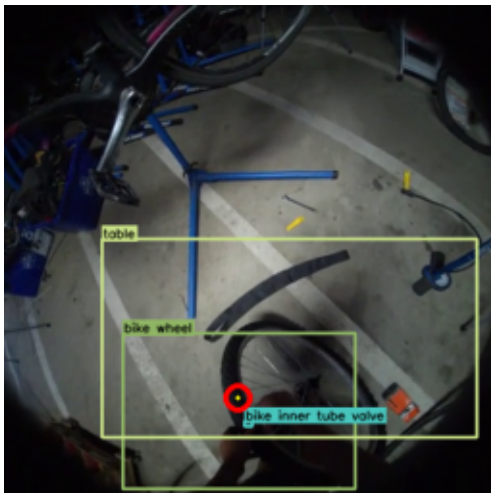
georgiatech_bike_06_10

Ego-Exo4D | Fix a Flat Tire - Replace a Bike Tube (real egocentric)

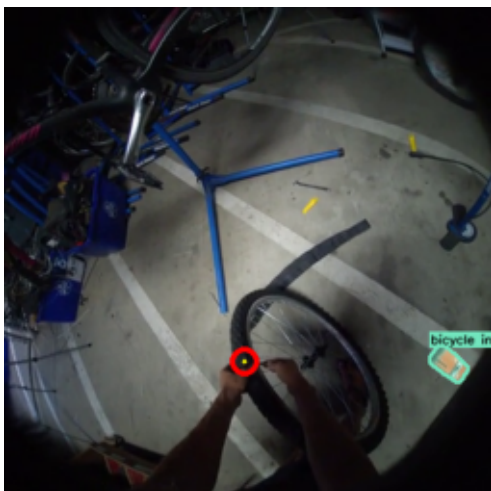
15 annotated steps + annotation inventory + methodology



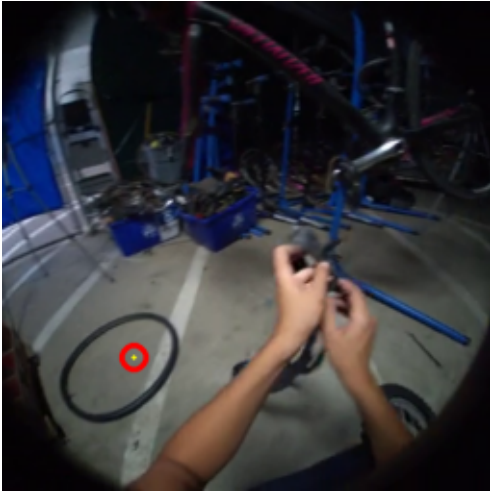
Step 1 [0-15s]: Unpack the new tube - Unpack the new tube from the carton (essential) | overlays: red=eye-gaze; boxes=bicycle inner tube pack, bike inner tube , left hand, right hand | atomic@~7s: "C holds a tube with her right hand."



Step 2 [15-27s]: Remove dust cap from the wheel - Remove the dust cap from the wheel (essential) | overlays: red=eye-gaze; boxes=bike inner tube valve, bike wheel, table | atomic@~21s: "C opens the tire valve with her right hand."



Step 3 [28-49s]: Squeeze out any air inside the tube - Deflate the tube to remove it from the tire (essential) | overlays: red=eye-gaze; masks=bicycle inner tube pack



Step 4 [88-92s]: Get a new inner tube - Get the new inner tube from the floor (essential) | overlays: red=eye-gaze
| atomic@~90s: "C holds a tube with her right hand."



Step 5 [92-96s]: Remove the valve cap - Unscrew the valve cap from the new tube (essential) | overlays: red=eye-gaze; masks=bicycle inner tube pack, bike inner tube ; boxes=bike inner tube valve | atomic@~94s: "C places a valve cap on the ground with her right hand."



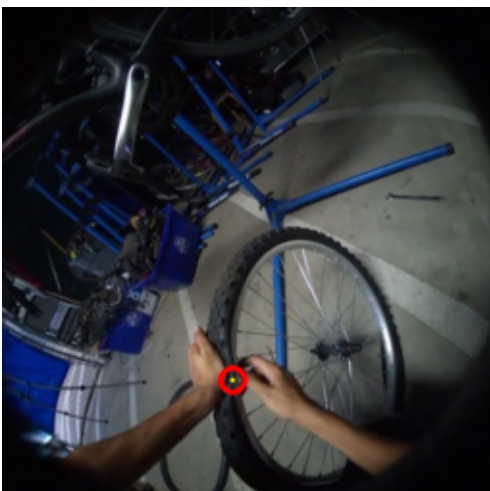
Step 6 [97-117s]: Engage the valve stem into the rim - Fix valve of new tube into the rim (essential) | overlays: red=eye-gaze; masks=bike inner tube | atomic@~107s: "C puts a tube in the tire with her right hand."



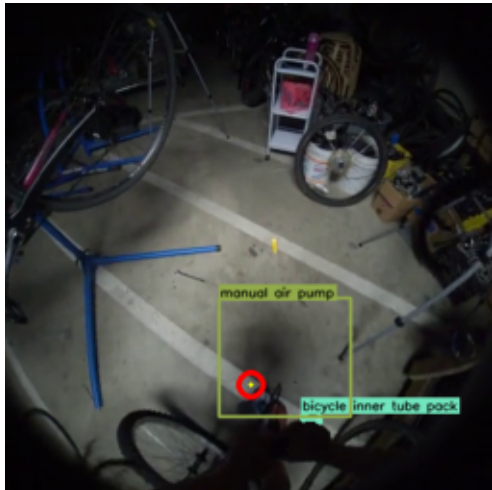
Step 7 [117-160s]: Fit the tube to the wheel - Fix the new tube to the wheel (essential) | overlays: red=eye-gaze
| atomic@~139s: "C holds the tire with her right hand."



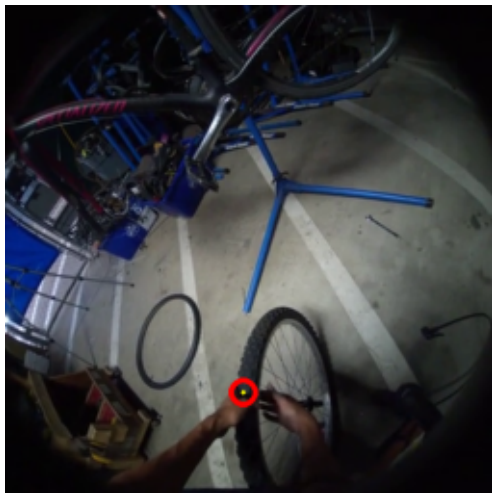
Step 8 [160-186s]: Check for any damage or splits in the thread - Check the tire thread for any damage or split (essential) | overlays: red=eye-gaze; masks=bike inner tube
| atomic@~173s: "C turns the tire with both hands."



Step 9 [186-197s]: Engage the valve stem into the rim - Fix the valve of new tube fully into the rim | overlays: red=eye-gaze
| atomic@~192s: "C holds the spokes with her right hand."



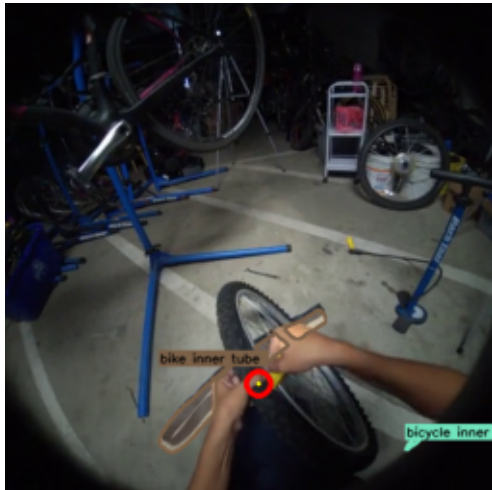
Step 10 [198-258s]: Fully inflate the tire - Inflate the tire with air pump (essential) | overlays: red=eye-gaze; masks=bicycle inner tube pack; boxes=manual air pump



Step 11 [259-273s]: Tighten the Dust Cap - Tightly screw the cap of the valve (essential) | overlays: red=eye-gaze | atomic@~266s: "C moves the air pump with her right hand."



Step 12 [51-53s]: Get bike tire lever - Get tire lever from the floor (essential) | overlays: red=eye-gaze; boxes=yellow tire lever | atomic@~52s: "C picks a tire lever from the floor with her right hand."



Step 13 [53-60s]: Engage a tire lever anywhere on the rim except right at the valve - Engage tire lever on the rim (essential) | overlays: red=eye-gaze; masks=bicycle inner tube pack, bike inner tube | atomic@~56s: "C engages the tire with the tire lever in her right hand."



Step 14 [60-81s]: Pull back and lift the bead out of the rim - Pull back and lift the bead out of the rim (essential) | overlays: red=eye-gaze | atomic@~70s: "C holds the tire with her right hand."



Step 15 [81-88s]: Pull the inner tube out - Pull the old inner tube out of the tire (essential) | overlays: red=eye-gaze; masks=bicycle inner tube pack; boxes=bike inner tube | atomic@~85s: "C holds a tube with her left hand."

ANNOTATION INVENTORY -- georgiatech_bike_06_10 (Fix a Flat Tire - Replace a Bike Tube)

take: 273s, 8188 frames @ 30fps ; ego RGB aria05_214-1.mp4 (448x448) ; scenario 'Fix a Flat Tire - Replace a Bike Tube'.

1. Eye-gaze 2729 per-frame 2D samples (also 3D + summary.json) -> red dot
[SPATIAL, sensor]
2. Keystep subgoals 15 segments (name+desc+essential) -> caption [TEMPORAL, human]
3. Relations - boxes 11 tracks (objects + L/R hands), 11 box annots -> labelled rectangles
[SPATIAL, human]
4. Relations - masks 5 objects masked (~1Hz COCO-RLE) -> filled+outlined [SPATIAL, human]
5. Atomic descriptions ... 151 timestamped action sentences -> caption / video subtitle
[TEMPORAL, human]
6. Expert commentary 6 spoken tip/error entries [TEMPORAL, expert]
7. Proficiency demo not annotated for this take [take-level, expert]

(Fields present in Ego-Exo4D but not spatial / not in the CoF focus scope:

procedure_understanding,

ego_pose, proficiency_demonstrator, physical_setting, visual_objects, participants -- listed for completeness.)

How these annotations were produced (Ego-Exo4D) -- ALL are real human/sensor labels on real video:

- Eye-gaze (red dot): Aria glasses' inward eye-tracking cameras; machine-perception computes per-frame 3D

gaze and projects it into the RGB camera (~10 Hz). The wearer's ACTUAL gaze, not a manual label.

- Keystep / subgoals: annotators segment the take into fine procedural steps from a per-scenario taxonomy;

each = start/end time + step name + description + 'essential' flag.

- Relations -- object/hand MASKS (filled + outlined) and BOXES (labelled rectangles): annotators use an

interactive point-click segmentation tool on SAMPLED frames (masks ~1 Hz) to mask/track task-relevant

objects and both hands; the same object is corresponded across ego<->exo views. Masks are stored as

LZString-compressed COCO-RLE; only a subset of objects is masked per take (varies 2-11 here).

- Atomic descriptions: narrators write timestamped fine-grained action sentences (subject + ego-visible flag).

- Expert commentary: a domain expert records spoken tips/error notes at timestamps.

- Proficiency demonstration (some takes): expert lists good executions + tips-for-improvement.