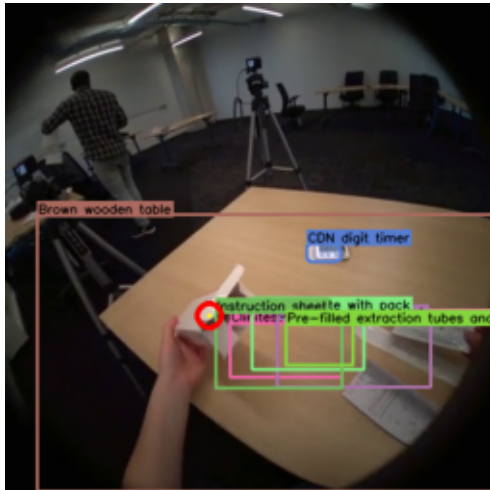


georgiatech_covid_02_10

Ego-Exo4D | Covid-19 Rapid Antigen Test (real egocentric)

16 annotated steps + annotation inventory + methodology



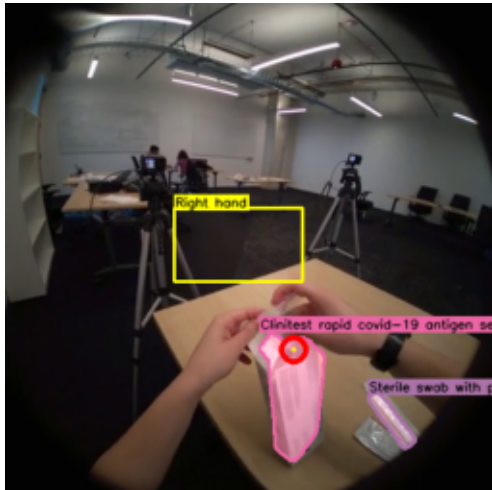
Step 1 [1-25s]: Arrange test material - Arrange test materials (essential) | overlays: red=eye-gaze; masks=CDN digit timer; boxes=Brown wooden table, Clinitest rapid covid-19 antigen self-test kit with pack, Instruction sheet, Pre-filled extraction tubes and dropper tips with pack, Sterile swab with pack, Test cassette with pack | atomic@~13s: "C places the test strip sachet on the table with her right hand."



Step 2 [26-35s]: Read the instructions - Open the instructions sheet (essential) | overlays: red=eye-gaze; masks=Test cassette with pack | atomic@~30s: "C spreads the test manual on the table with both hands."



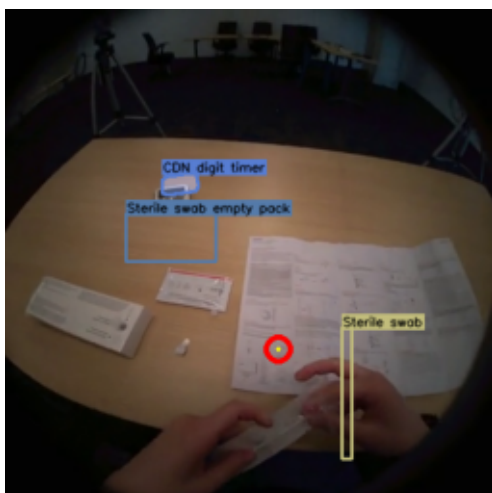
Step 3 [35-51s]: Arrange test material - Arrange test materials (essential) | overlays: red=eye-gaze; masks=Clinitest rapid covid-19 antigen self-test kit with pack | atomic@~43s: "C places the extraction tube sachet on the table with her left hand."



Step 4 [51-56s]: return the test kit items back into the box - Put the pack of dropper into the test kit | overlays: red=eye-gaze; masks=Clinitest rapid covid-19 antigen self-test kit with pack, Sterile swab with pack; boxes=Right hand | atomic@~53s: "C returns the extraction tube sachet into the test box with her right hand."



Step 5 [56-71s]: Read the instructions - Open the cap of the dropper and place it on the test kit | overlays: red=eye-gaze; masks=CDN digit timer; boxes=Pre-filled extraction tube, Pre-filled extraction tube lid | atomic@~64s: "C opens the extraction tube lid with her right hand."



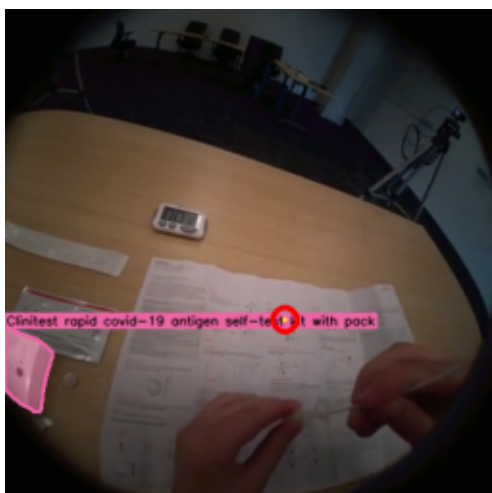
Step 6 [72-89s]: Locate and unwrap the collection swab - Remove the swab from the wrap and set the timer (essential) | overlays: red=eye-gaze; masks=CDN digit timer; boxes=Sterile swab, Sterile swab empty pack | atomic@~80s: "C passes the nasal swab sachet to her right hand."



Step 7 [89-92s]: Slowly insert the tip of the swab into the nose - Insert the swab into the nose (essential) | overlays: red=eye-gaze; masks=Clinitest rapid covid-19 antigen self-test kit with pack | atomic@~91s: "C swabs her right nostrils with the swab in her right hand."



Step 8 [92-127s]: Rotate and swirl the swab around for 5-10 times - Rotate and swirl the swab (essential) | overlays: red=eye-gaze; boxes=CDN digit timer | atomic@~110s: "C swabs her left nostrils with the swab in her right hand."



Step 9 [131-144s]: Dip the swab into the testing tube - Dip the swab into the testing tube and swirl (essential) | overlays: red=eye-gaze; masks=Clinitest rapid covid-19 antigen self-test kit with pack | atomic@~138s: "C rolls the nasal swab inside the extraction tube with her right hand."



Step 10 [145-150s]: Set a timer - Set the timer for the test (essential) | overlays: red=eye-gaze; masks=Sterile swab, Sterile swab empty pack, Test cassette with pack | atomic@~147s: "C sets the timer with both hands."



Step 11 [150-172s]: Read the instructions - Read the instructions (essential) | overlays: red=eye-gaze



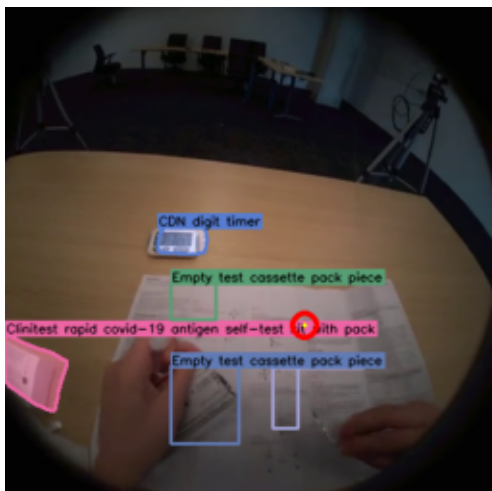
Step 12 [172-207s]: Wait for the testing solution incubation time - Wait for the test result (essential) | overlays: red=eye-gaze; masks=Instruction sheet, Sterile swab empty pack



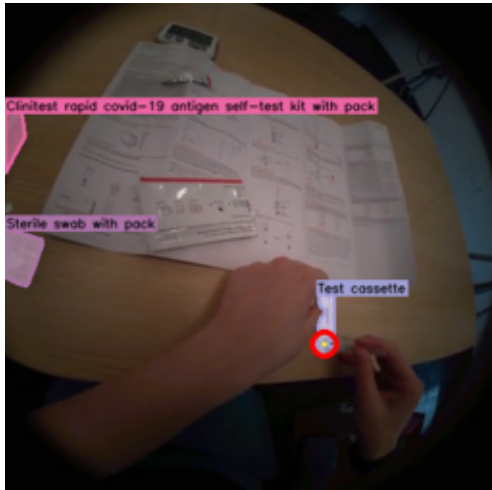
Step 13 [207-221s]: Rotate and swirl the swab around for 5-10 times - Press the testing tube (essential) | overlays: red=eye-gaze; masks=Clinitest rapid covid-19 antigen self-test kit with pack | atomic@~214s: "C squeezes the extraction tube with her left hand."



Step 14 [227-230s]: cover the test tube - Cover the test tube with a lid (essential) | overlays: red=eye-gaze | atomic@~228s: "C covers the extraction tube with the cap in her right hand."



Step 15 [230-241s]: Unwrap testing plate - Remove the testing plate from the foil pack (essential) | overlays: red=eye-gaze; masks=CDN digit timer, Clinitest rapid covid-19 antigen self-test kit with pack; boxes=Empty test cassette pack piece, Test cassette | atomic@~235s: "C drops the torn sachet on the table with her right hand."



Step 16 [243-254s]: Apply a few drops of the testing solution to the testing plate - Pour few drops of testing solution into the testing plate (essential) | overlays: red=eye-gaze; masks=Clinitest rapid covid-19 antigen self-test kit with pack, Sterile swab with pack, Test cassette | atomic@~248s: "C drops the fluid in the receiving port of the test strip with the extraction tube in her right hand."

ANNOTATION INVENTORY -- georgiatech_covid_02_10 (Covid-19 Rapid Antigen Test)

take: 264s, 7912 frames @ 30fps ; ego RGB aria01_214-1.mp4 (448x448) ; scenario 'Covid-19 Rapid Antigen Test'.

1. Eye-gaze 2637 per-frame 2D samples (also 3D + summary.json) -> red dot
[SPATIAL, sensor]
2. Keystep subgoals 16 segments (name+desc+essential) -> caption [TEMPORAL, human]
3. Relations - boxes 16 tracks (objects + L/R hands), 16 box annots -> labelled rectangles
[SPATIAL, human]
4. Relations - masks 11 objects masked (~1Hz COCO-RLE) -> filled+outlined [SPATIAL, human]
5. Atomic descriptions ... 93 timestamped action sentences -> caption / video subtitle
[TEMPORAL, human]
6. Expert commentary 5 spoken tip/error entries [TEMPORAL, expert]
7. Proficiency demo present: 18 good-exec + 22 tips [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.