WORTEL

EGOCENTRIC VIDEO ANNOTATION

Segmentation Rules and Guidance

How to cut first-person video into atomic-action clips. Follow these rules for every clip. Reviewers check every clip against them.

Six rules to remember

1. One clip = one action.
2. Split when the hand, the action or the object changes.
3. Keep one continuous action in one clip.
4. The hand must be visible at the start and end of every clip.
5. No clip is longer than 10 seconds.
6. Clips never overlap.

1 One clip = one action

A clip shows one complete action, from its start to its end.

Quick test: can you describe the clip in one short phrase, like "pick up the cup"? If you need two phrases, make two clips.

WRONG three actions in one clip

- Pick up cup → walk to table → put down cup

CORRECT three clips

- Clip 1: Pick up cup
- Clip 2: Walk while holding cup
- Clip 3: Put down cup

These are always separate clips: pick up, hold, carry, put down, open, close, turn on, turn off.

2 When to split

Start a new clip when the hand, the action or the object changes.

Ask these four questions. If any answer is yes, start a new clip.

1. Did the hand change? (left to right, right to left, or one hand to both hands)

WRONG one clip

- Left hand puts down cup → right hand picks up cloth

CORRECT two clips

- Clip 1: Put down cup (left hand)
- Clip 2: Pick up cloth (right hand)

2. Did the action change?

WRONG one clip

- Wash vegetables → turn off faucet

CORRECT two clips

- Clip 1: Wash vegetables
- Clip 2: Turn off faucet

The same rule applies here:

- Open cabinet door → take out bowl = 2 clips
- Right hand applies tape → smooths tape = 2 clips

3. Did the object change?

WRONG one clip

- Pick up cup → pick up cloth

CORRECT two clips

- Clip 1: Pick up cup
- Clip 2: Pick up cloth

The same action on different items = 1 clip for each item, even if the items are the same type.

4. Do different steps repeat in a cycle?

WRONG one clip

- Cut → drop in basket → cut → drop in basket

CORRECT one clip for each step

- Clip 1: Cut
- Clip 2: Drop in basket
- Clip 3: Cut
- Clip 4: Drop in basket

More examples

Pick up vegetables and wash them:

WRONG one clip

- Pick up vegetables → put in sink → start washing

CORRECT three clips

- Clip 1: Pick up vegetables
- Clip 2: Put vegetables in sink
- Clip 3: Wash vegetables

Cloth in a basin:

CORRECT three clips

- Clip 1: Rinse cloth in basin
- Clip 2: Pick up and wring cloth
- Clip 3: Carry cloth away

Pick up, inspect, put back:

CORRECT three clips

- Clip 1: Pick up item
- Clip 2: Inspect item
- Clip 3: Put item back

Clothes: fold, straighten, iron, adjust collar, button = one clip for each. If an item is picked up and put straight back in one quick motion, 1 clip is fine.

3 When to keep one clip

Do not split one continuous action.

Same hand, same action, same object, no stop

WRONG a new clip for every wipe

- Wipe → wipe → wipe → wipe

CORRECT one clip

- Wiping the table continuously

The same applies to cutting vegetables, washing, sweeping and mopping. Split only for the 10-second limit.

Two hands working together at the same time

ONE CLIP hands work together

- Left hand opens cabinet + right hand takes bowl
- Left hand holds an object + right hand works on it

TWO CLIPS one hand after the other

- Left hand puts down cup → pause → right hand picks up cloth

Small movements

Finger movements and grip changes belong to the main action.

WRONG three clips

- Move hand to cup → touch cup → pick up cup

CORRECT one clip

- Pick up cup

WRONG a clip for each finger movement

- Move finger → adjust finger → move finger → adjust finger

CORRECT one clip

- Pinch the bun

Making a bun = four clips only: pick up wrapper, add filling, pinch bun, put down bun.

4 Hand visibility

At the start and end of every clip, the working hand must be visible from wrist to fingers. For a two-hand action, one visible hand is enough.

WRONG

- The clip starts while the hand is hidden.
- The clip ends while the hand is hidden.
- Only the arm is visible, not the wrist and hand.

CORRECT

- The wrist and hand are visible in the first frame and in the last frame.
- Picking up a bottle: the clip shows the hand before it grabs the bottle and after it is holding it.

During the clip, the hand should be visible at least 90% of the time.

DO NOT MAKE A CLIP WHEN

- The hand is hidden for 1 second or more without a break.
- The hand is hidden for more than 20% of the clip.
- You cannot tell which hand is working.
- The hand is behind glass or a mirror.
- It is too dark to see the hand.

EXAMPLES

- Hand goes into a sink, cabinet, drawer or shelf for a short moment and the action clearly continues = keep the same clip.
- Hand is about to leave the frame = end the clip before it leaves.
- Hand keeps going in and out of view = clip only the visible parts. Do not join them into one clip.
- Hand disappears inside a drawer for 2 seconds = no clip for that part.

5 Clip length and timing

Maximum 10 seconds. Split long actions at least every 10 seconds, at a natural pause if there is one.

WRONG one clip

- Wiping the table for 25 seconds

CORRECT three clips, each 10 seconds or less

- Clip 1: Wipe table
- Clip 2: Wipe table
- Clip 3: Wipe table

No overlaps. A red line on the timeline means two clips overlap. Move the clip edge with A/S or D/F until the red line is gone.

WRONG

- Clip 1 ends at 00:04.12 and Clip 2 starts at 00:04.11.

CORRECT

- Clip 1 ends at 00:04.12 and Clip 2 starts at 00:04.13 or later.

Hand label. Choose the hand that does the action: left, right or both. Left marked as right, or one hand marked as both, is a fatal error.

Very fast actions. An action under 0.5 seconds, such as quickly pressing a switch or quickly turning off a faucet, can join the action before or after it.

Exact start and end. Start the clip where the action starts. End it where the action ends.

Camera shake. End the clip before the shake starts.

6 What to record

ACTIONS TO RECORD	
Pick up an object, including food and drink	YES
Put down an object, including food and drink	YES
Hold an object	YES
Carry an object while walking	YES
Open or close a door, cabinet or drawer	YES
Turn a faucet or switch on or off	YES
Take an object out of a cabinet, shelf, tool bag or pouch, or put it in	YES
Wash, cut, wipe, mop, sweep, rinse, wring, fold or tape an object	YES
Touch an object on purpose	YES
Apply soap to hands	YES
Wash hands with soap	YES
Dry hands with a towel	YES

ACTIONS NOT TO RECORD	
Walking with empty hands	NO
Empty or idle hands	NO
Eating or drinking	NO
Smelling an item	NO
Using a phone, laptop or touchscreen	NO
Using a mouse or keyboard while sitting at a computer	NO
Washing your own hands with water only, no soap or towel	NO
Scratching	NO
Taking things from clothing pockets, or putting them in	NO
Putting on gloves	NO
Adjusting gloves that are already on	NO
Actions by another person	NO
Smoking, or handling a gun	NO
Hand movements behind glass or a mirror	NO
Small finger movements that are part of a bigger action	NO

Example: drinking

Pick up cup → drink → put down cup

TWO CLIPS

- Clip 1: Pick up cup
- No clip: Drinking
- Clip 2: Put down cup

Example: washing hands

Turn on faucet → pick up soap → wash hands with soap → put down soap → turn off faucet

FIVE CLIPS

- Clip 1: Turn on faucet
- Clip 2: Pick up soap
- Clip 3: Wash hands with soap
- Clip 4: Put down soap
- Clip 5: Turn off faucet

Drying hands with a towel is also recorded. Rinsing hands with water only, with no soap or towel, is not.

Example: walking

RECORD

- Walking while holding a cup.

DO NOT RECORD

- Walking with empty hands. Leave that part of the timeline blank.

Example: computer

Sitting at a computer

NO CLIPS

- Using the mouse or keyboard.
- Typing, scrolling or tapping on a phone or laptop.

7 Task Error

Use Task Error only when the **whole video** cannot be annotated. If only one part is bad, annotate the good part.

MARK TASK ERROR WHEN THE WHOLE VIDEO HAS ONE OF THESE

- No hands, or almost no hands.
- Too dark, fully black or fully white.
- Ripples, flickering, or green, red or black bars and lines.
- Mostly phone use.
- Not a first-person view.
- The wearer has black gloves on.
- The camera wearer's face is visible.

ACCEPTABLE, SO NO TASK ERROR

- Faint black bars.
- Only the wearer's nose is visible.

WRONG

- One part of the video is dark and the rest is clear → Task Error

CORRECT

- One part of the video is dark and the rest is clear → annotate the clear part
- No hands visible in the whole video → Task Error
- Green or black flickering through the whole video → Task Error

8 Fatal errors

These four errors get the work rejected. An annotator who makes them three times is removed from the project.

1. More than one action in a clip

- One clip holds the end of "turn on faucet" and the start of "pick up soap". It must be two clips.

2. Wrong hand label

- Left hand marked as right.
- One hand marked as both.

3. Hand not visible at the start or end

- The first or last frame does not show the wrist and hand.

4. Overlapping clips

- A clip starts before the previous clip ends (red line on the timeline).

Reviewers also mark these as serious errors: a clip over 10 seconds, over-splitting, a missed action, a clip for something on the "do not record" list, and a wrong Task Error.

9 Check before you submit

- [ ] One action in every clip.
- [ ] No action missed.
- [ ] Nothing from the "do not record" list.
- [ ] Hand visible at the start and end.
- [ ] No clip over 10 seconds.
- [ ] No overlaps.
- [ ] Hand labels correct.
- [ ] No over-splitting.
- [ ] Task Error only for a fully unusable video.

10 For reviewers

1. Check every task and every clip, not only the flagged ones.
2. Scrub the full timeline first to find missed actions.
3. For each clip, check: valid action, one action, start and end, hand label, visibility, length, overlap.
4. Write down each error: the task, the clip and the rule broken.
5. Look for mistakes that repeat across the account.
6. Decide. **Approved:** no mistakes. **Rework:** mistakes found; send the error list, and the annotator rechecks every task.
7. After rework, check the fixed clips first, then the rest of the account again.

11 Shortcuts

KEY	WHAT IT DOES	KEY	WHAT IT DOES
N	Start a clip. Press again to end it.	**A / S**	Move the start.
Space	Play or pause.	**D / F**	Move the end.
L	Loop the clip.	**B**	Split at the playhead.
X	Delete the selected clip.	**Ctrl + Z**	Undo.