HRI: Off-line Detection and Tracking Tests

version 1.0/12-Aug-2025/ZH

Purpose

This test is designed to evaluate a human-detection and object tracking pipeline off-line, with pre-recorded images and ground-truth boxes, to measure detection quality and tracking consistency. This is intended as a **pre-test** procedure that must be demonstrated successfully before advancing to the **HRI: Live Detection and Tracking Test** and the **Safety: Stopping and Collision Avoidance Test**. See Figure 1 for an explanation of the permissible test sequences.

Test Facility

This test comprises an off-line evaluation using a data set from the public **ARRnet Tracking Test** repository (see GitHub page [1]). No field site or physical robot is required. The results of this test will determine if a robot system can advance to further, more complex and higher risk tests. A passing score for this test will enable the robot system to engage in further tests with human participants.

This test uses image sequences and accompanying ground-truth human annotations from the repository [1]. The primary objective of the task is to detect people within the images, and the secondary, optional, objective is to maintain people's identity (anonymised, i.e. labelled within the data set) over time. Metrics evaluate *detection quality* (Precision/Recall/F1) and *tracking stability* (ID-switches).

For technical integration details and ROS message specifications, refer to the repository README [1].

Test environment and configuration

- Compute/Runtime Requirements: ROS 2 Humble on Ubuntu 22.04 (or equivalent), CPU/GPU as per detector requirements. Ensure consistent hardware across runs.
- ROS 2 workspace with `tracking_test` package built via `colcon`.
- Detector node: implemented via `detector_node.py` (`mock_detection_publisher.py` is for demonstration only and not for official testing).
- Visualisation: `visualiser.py` (publishes `/eval_viz`), optional `rqt_image_view`.

Test Procedure

Run all provided images sequences [1] from beginning to end. The labelling policy is to treat ground-truth “human” annotations as positives. Non-human classes are out of scope for scoring.

1. Detection-Only Test

- Objective: To evaluate detection performance regardless of tracking.
- Procedure: Run the evaluator with your detector; record per-frame and final metrics.
- Outputs: TP, FP, FN, Precision (P), Recall (R), optional F1.

2. Tracking Consistency Test (optional)

- Objective: To identify stability over time.
- Procedure: Enable your tracker; run the same sequences.
- Outputs: ID-switch count (per sequence and per 100 frames), min/mean track length.

Evaluation Criteria and Error Conditions

Trials are evaluated according to the following metrics, computed using the off-line data set provided:

- **Precision (P)**, **Recall (R)** and **F-measure (F1)** computed from number of True Positives (TP), False Positives (FP) and False Negatives (FN):

  P = TP / (TP + FP)

  R = TP / (TP + FN)

  F1 = (2 * TP) / ((2 * TP) + FP + FN) = 2 * (**P** * **R**) / (**P** + **R**)

- **ID-switches** (count) — and normalised per 100 frames.
- **IoU threshold** used for matching (document value; default IoU ≥ 0.5 unless configured otherwise).
- **Confidence threshold** used by your detector (document value).

Note: This procedure does not enforce numeric pass/fail thresholds. Define targets per project (e.g., P≥0.9, R≥0.9, ID-switches ≤ X/100 frames).

Conditions

- Use the repository dataset (tracking_test/dataset).
- Keep detector and tracker configs constant during comparisons.
- Time synchronisation: evaluator aligns topics by timestamp (50 ms slop). Ensure your detector timestamps outputs correctly.

Test Result

The following shall be recorded and included in the test report:

- Sequence name, frame count, detector config (model, input size, confidence/NMS thresholds).
- TP/FP/FN counts; P, R, F1; IoU threshold used.
- ID-switch count and rate.
- Annotated video/`images` from `/eval_viz.`

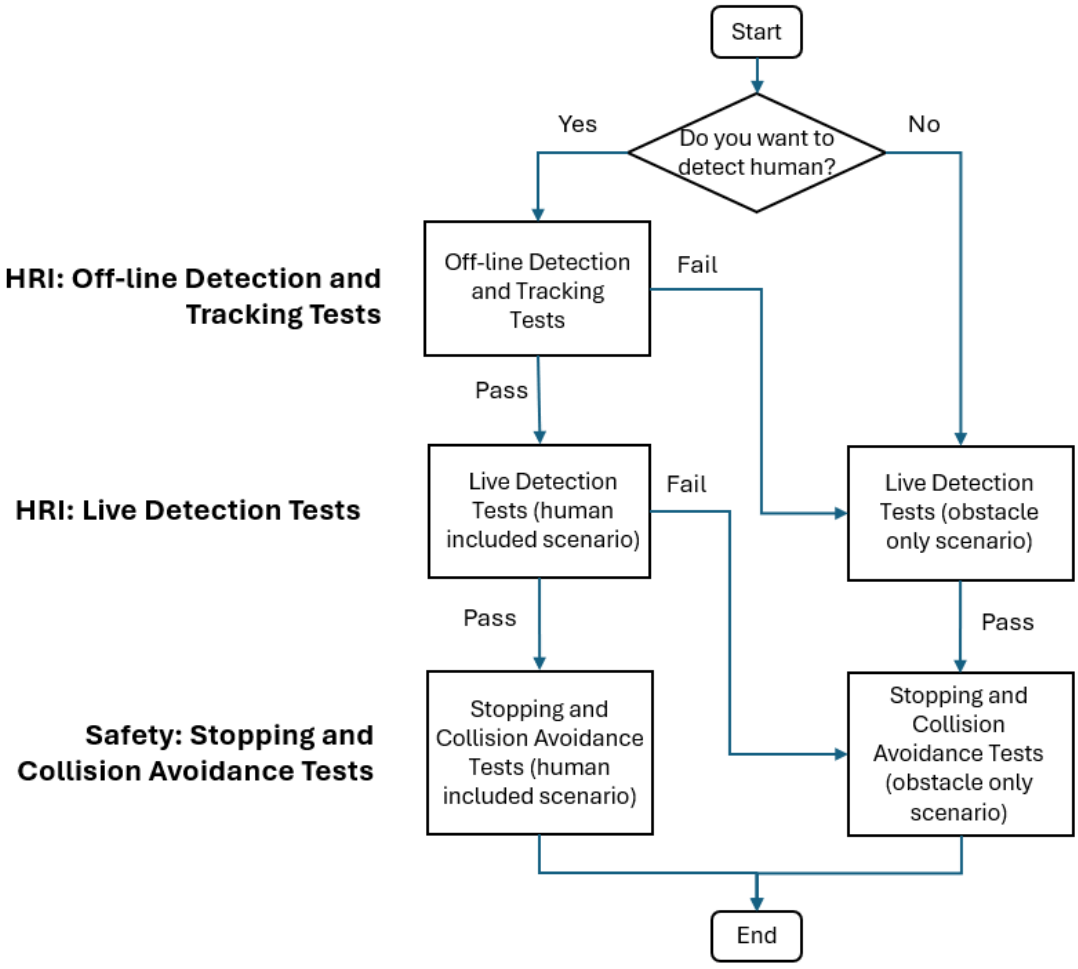

Figure 1. Test Sequence Flow

References

1. https://github.com/LCAS/ARRnet_tracking_test