University of Nottingham
UK | CHINA | MALAYSIA

Department of Electrical and Electronic Engineering

Robot-Elevator Interaction:
From Physical Panel Recognition to Activation

NGIOE Next Generation Internet of Everything Laboratory

1 Motivation and problem

Service robots need access to existing elevators without modifying their controls. Panels vary in layout, labels and lighting. The challenge is to identify the requested button, keep its identity stable as the view changes, and turn that target into a controlled press.

2 Hardware platform

01 WRIST RGB CAMERA
TYS-CLB30W-V1
Capture: 640 × 480 (0.31 MP), 30 fps

02 BAE
Button Actuation End-effector

03 ROBOTIC ARM
Waveshare RoArm-M3
5 joints + gripper | 360° base rotation
Rated payload: 0.2 kg at 0.5 m

04 ONBOARD COMPUTER
NVIDIA Jetson AGX Orin
64GB Developer Kit

05 POWER STATION
FG300B+ | Battery: 299.7 Wh
Max. total output: 300 W



06 RGB-D CAMERA
Orbbec Gemini 335
RGB: up to 1920 × 1080 @ 30 fps
Depth: up to 1280 × 800 @ 30 fps
Depth FOV: 90° H × 65° V

07 LAPTOP DISPLAY
AGX Orin remote display
NoMachine (NX): remote desktop
SSH: terminal access

08 MOBILE TEST CART
Three-tier utility cart
Nominal size (L × W × H):
810 × 365 × 750 mm
Rated total load: 200 kg

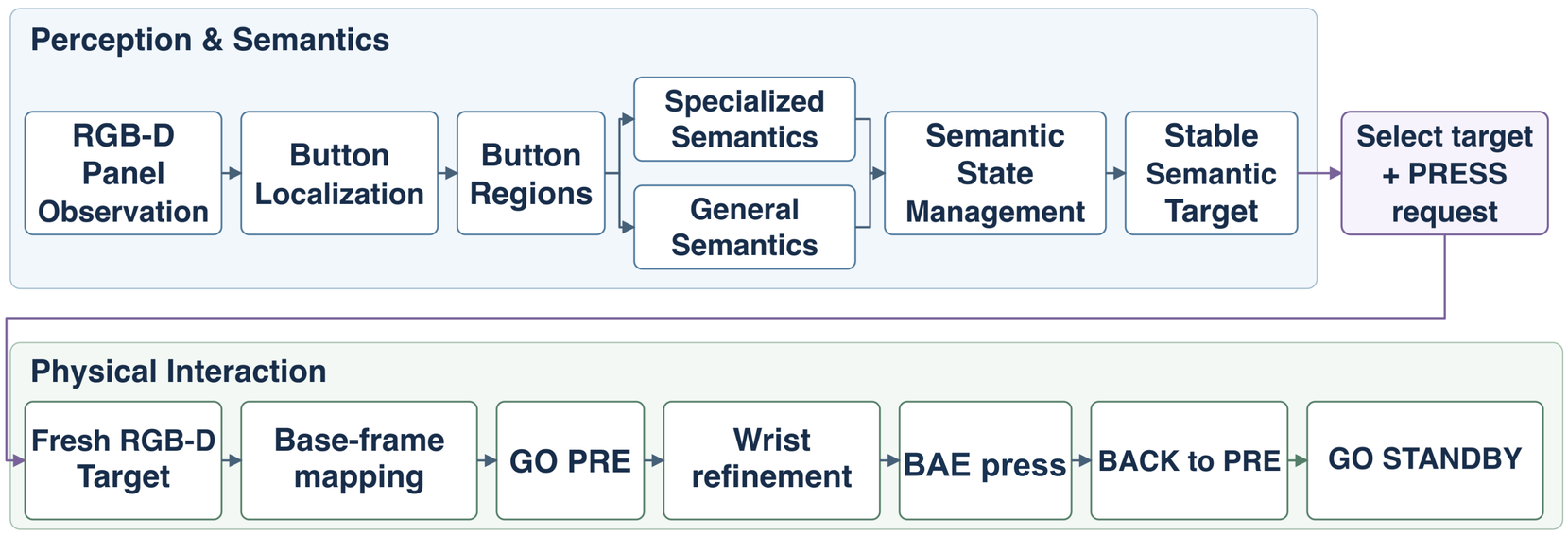
Hardware roles

RGB-D camera: panel view and 3D location.
Wrist camera: one alignment update.
Jetson AGX Orin: onboard computation.
Arm and BAE: press the selected button.

Operating scope

The operator positions the cart and selects the target.
During pressing, the cart stays still.

3 Software blocks



PRE = pre-contact pose. BAE = button-pressing tool. Standby = resting pose.

Reading the panel

Door controls: HOG+SVM reads Open / Close / Other.
Broader labels: Qwen reads floor numbers and Up / Down arrows.
State management retains confirmed labels when repeated readings agree.

Positioning the arm

Fresh RGB-D: refresh depth before motion.
Base-frame mapping: express position relative to the arm base.
Wrist refinement: adjust position using the wrist camera.

4 Recognition results

Saved-image detector tests.
Rounded rates per 100 (not sample counts).

Detector	Recall	Precision
YOLO11n Detect (button training)	**~98 / 100**	**~99 / 100**
Grounding DINO	~97 / 100	~55 / 100
YOLO-World	~72 / 100	~54 / 100

YOLO11n is trained for buttons. The other two use text prompts for open-vocabulary detection. Training conditions differ.

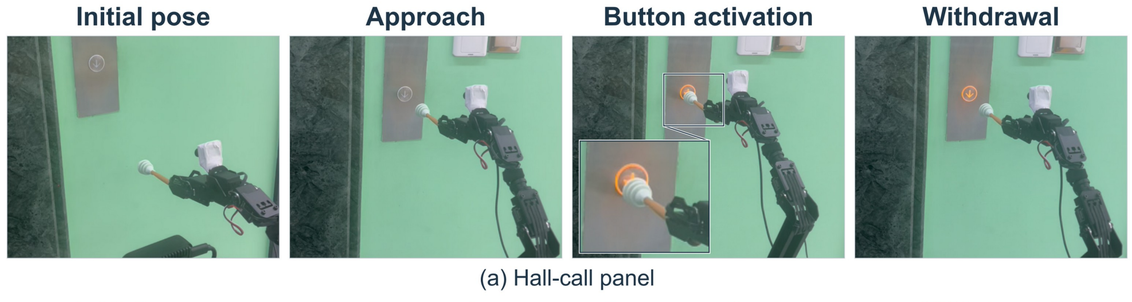
Recall: real buttons found / all real buttons.
Precision: correct boxes / all predicted boxes.
IoU: shared area / area covered by one or both boxes.
A match to a hand-labelled box requires IoU ≥ 0.50.

Measure	Result	Context
HOG+SVM Door-control macro-F1	**0.996**	379 button crops 0–1 score
Target selection (saved images)	**85.1%**	404 / 475 requests correct 112 saved RGB images
Unchanged label requests skipped	**99.8%**	3,532 / 3,540 requests in one 207-second run

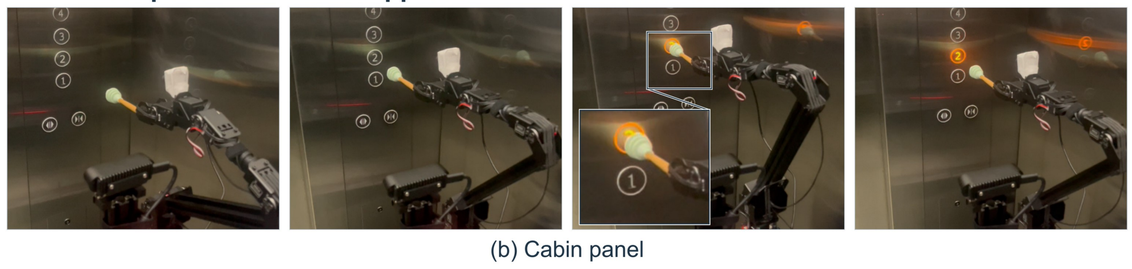
HOG = image-edge features. SVM classifies Open / Close / Other.
Macro-F1 averages class F1 scores. F1 balances recall and precision.
Skipping unchanged requests before inference avoids repeated work.
99.8% is a filtering rate, not recognition accuracy or press success.

5 Actuation results

Two PMB cases: hall-call (outside) and cabin floor 2 (inside).
Buttons lit during contact. The arm then withdrew.



(a) Hall-call panel



(b) Cabin panel

Evidence limits

Two activations do not establish repeated success rates or performance on other elevators. Video review checked the lights. Press-position accuracy and button force were not quantified.

6 Future work

Test repeated live perception-to-press cycles across elevators, viewpoints and lighting conditions. Record video with robot feedback to measure correct-target activation and repeatability. Investigate button-light tracking for online outcome checks and recovery.

Siheng Ye · Zejing Tong · Suiyi Zhang
Supervisor: David Chieng