Autonomous Mobile Manipulation in the Real World

from Torque Limits to Perception and Language

AI Robotics Lab

2026. 10. 30

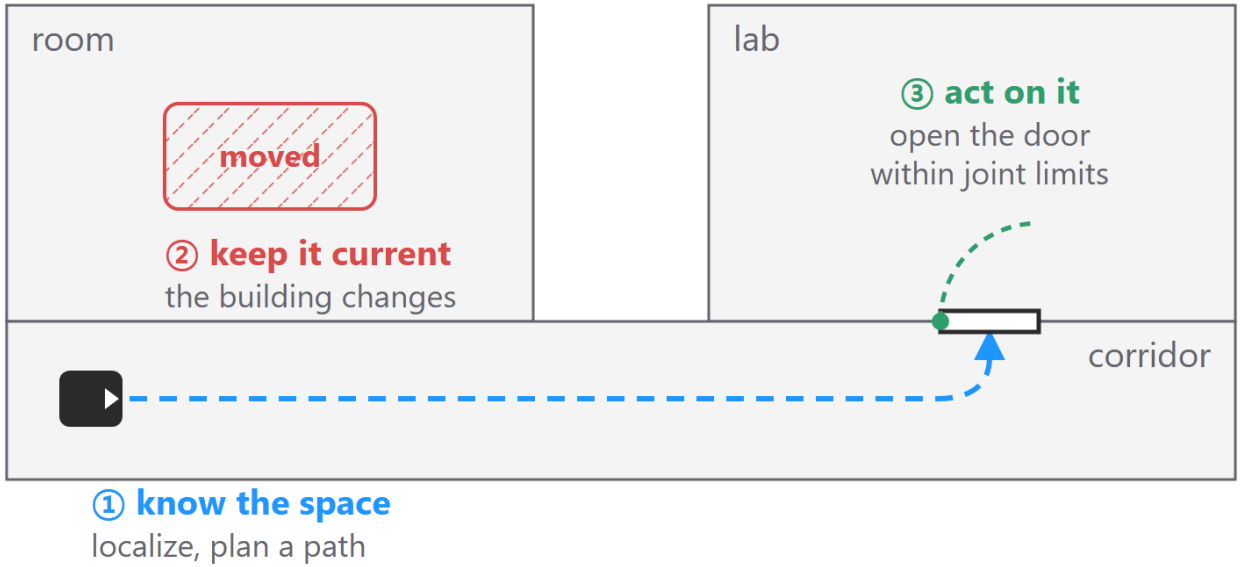
KangGeon Kim

AI Robotics Lab, Dept. of Electronic Engineering, Jeonbuk National University

Introduction

Robots in buildings: arriving is not the task

"A robot that only navigates stops at the first closed door"



① ② move the robot through the building. ③ changes the building.

Know the space
Localize in a map; plan where it can go.

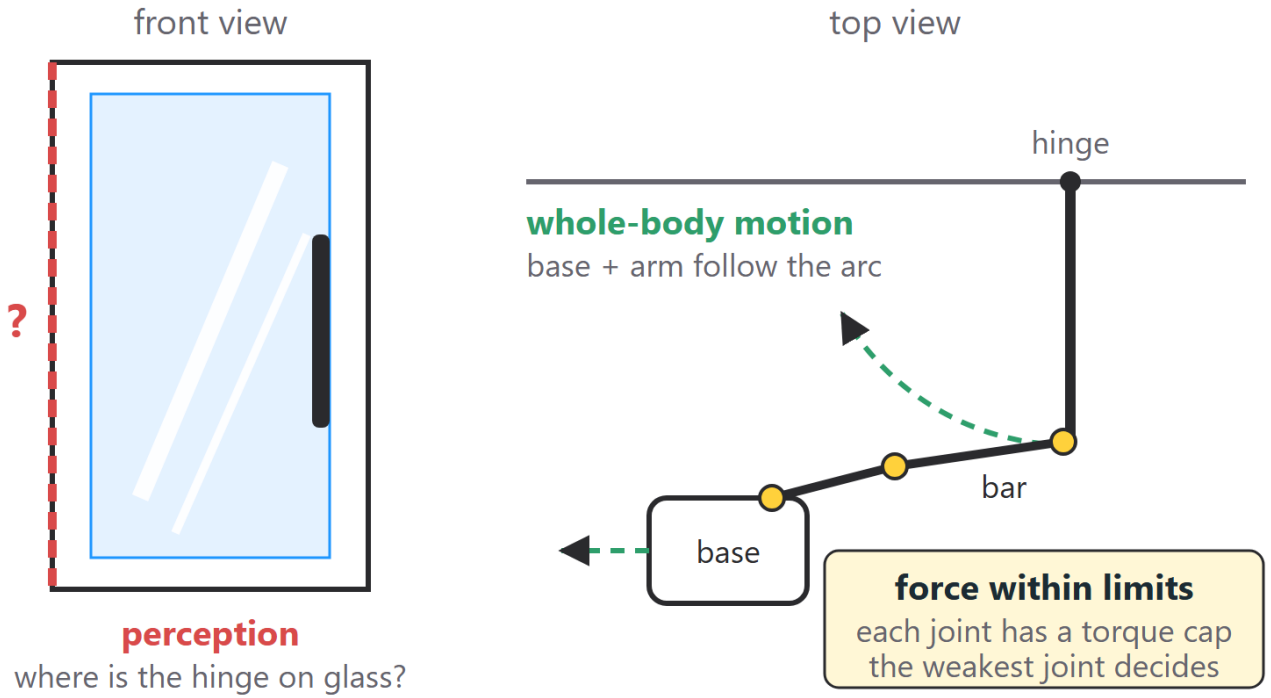
Keep it current
Buildings change: parked cars, moved furniture. The map must follow.

Act on it
Doors, drawers, elevators: contact, within the robot's own joint limits.

Introduction

The door: a small object that couples everything

"Opening a door takes perception, whole-body motion and force at once"



Perception
Where is the hinge? Glass and reflections make it hard.

Whole-body motion
Base and arm move together along a hinged arc.

Force within limits
A closer resists; every joint has a torque cap, and the weakest joint decides.

Roadmap

Three themes: measurements and a design

"Torque capability and perception were measured; the language pipeline is a design"

Act 3

Torque limits

Measured: how the start limits radial force capability.

→

Act 4

Perception

Measured: what a glass door lets us trust.

→

Act 5

Language

Designed: from exploration to a spoken goal.

Act 1: where this comes from. Act 2: a published planner meets a real robot.

1 Where this comes from

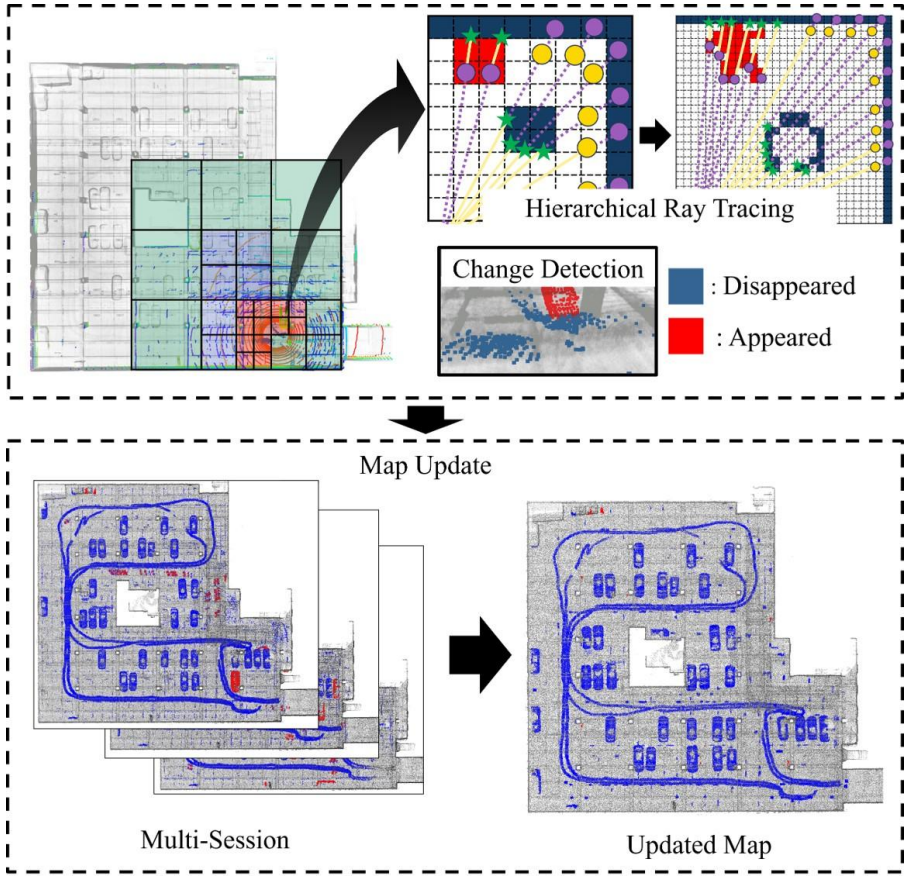
Maps that stay current — and still no contact

AI Robotics Lab

1. Where this comes from

Earlier work 1 · Fast change detection for 3D map update [1]

"Cast rays only where the map could have changed"



Problem

Maps go stale as vehicles come and go. Ray tracing finds changes, but wastes rays on unchanged regions and can falsely flag ceilings.

Method

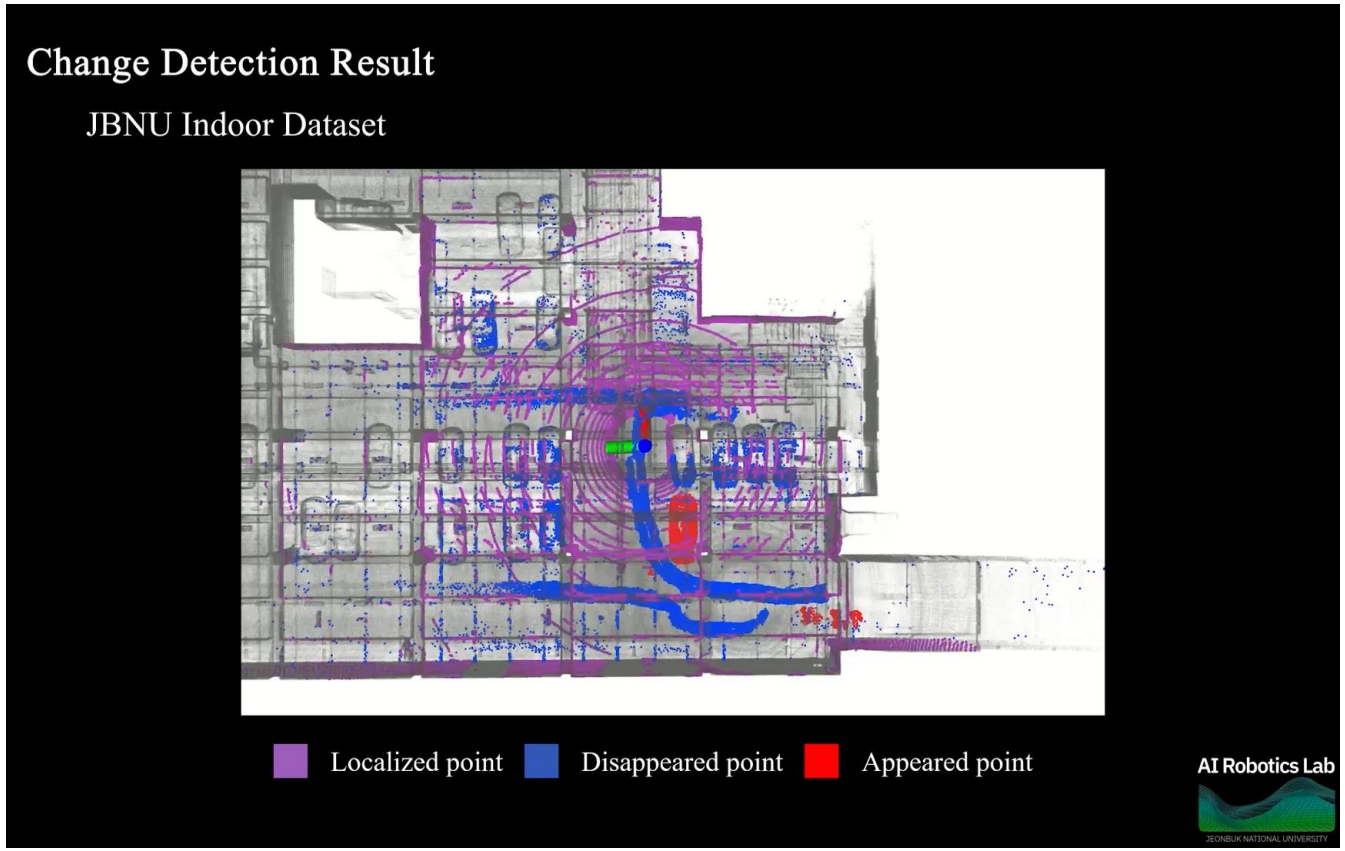
A coarse-to-fine hierarchy drops rays early. Consistency filtering and multi-session aggregation decide the update.

[1] K. Kim*, T. Jang*, J. Yoon, H. Jo, "Fast Change Detection of Point Clouds for 3D Map Update Using Hierarchical Structure," IEEE Sensors Journal, under review.

1. Where this comes from

Earlier work 1 · Results on indoor and outdoor parking lots [1]

"Fewer rays, fewer false positives, less time per frame"



0.161 → 0.847
Precision of map removal, indoor

×1.35 / ×1.13
Per-frame speedup, indoor / outdoor

0.370 → 0.797
Voxel IoU, indoor

Video: live change detection on the JBNU indoor lot, then LT-Parkinglot [2]. Blue = disappeared, red = appeared.

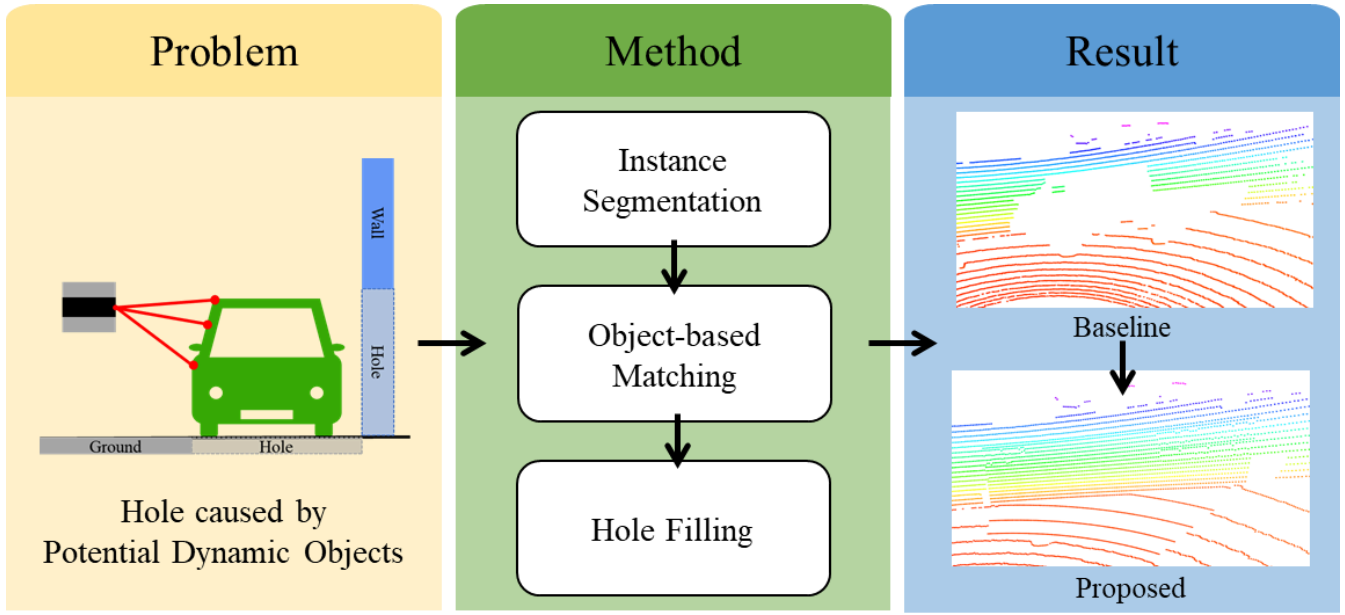
Recall stays high (0.970 vs 0.999).

[1] Kim et al., IEEE Sensors Journal, under review. [2] G. Kim, A. Kim, "LT-mapper: A modular framework for LiDAR-based lifelong mapping," ICRA 2022.

1. Where this comes from

Earlier work 2 · Removing potential dynamic objects and filling the holes [3]

"A parked car is not part of the building: remove it, then repair the map"



Problem

Parked vehicles sit still during mapping, so dynamic removal keeps them. Deleting them leaves holes in walls and ground.

Method

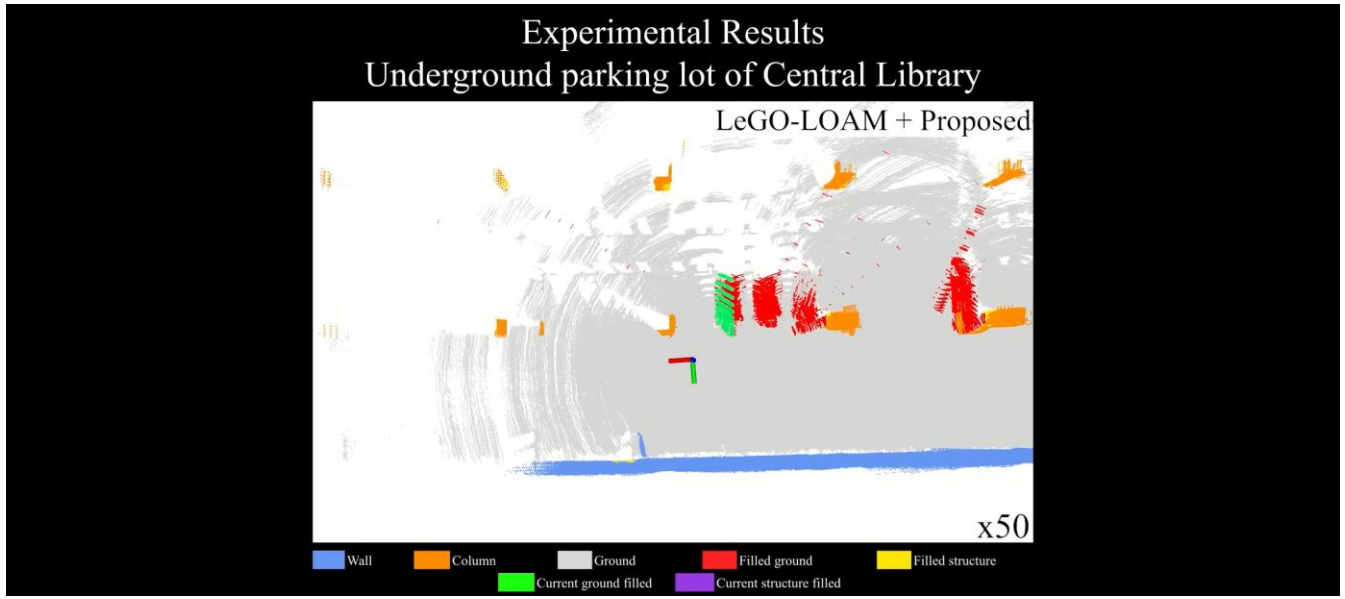
Voting-based instance segmentation and frame matching find them. Geometric hole filling restores ground and walls.

[3] K. Kim, H. Jo, "An Efficient Removal of Potential Dynamic Objects and Hole Filling Method for Static Point Cloud Maps," conference manuscript, under revision.

1. Where this comes from

Earlier work 2 · Results in two underground parking lots [3]

"Denser and more consistent static maps than removal alone"



75.6 → 97.9 %
F-score, Library 1 (removal only → + proposed)

0.072 → 0.030 m
Chamfer L1, Library 1

68.9 → 94.4 %
F-score, Center 1 (no retraining)

Video: map building with removal and hole filling, Library lot then International Center lot. Red = filled ground, yellow = filled structure.

Backends: LeGO-LOAM [4], Faster-LIO [5].

Both projects improved maps for navigation. Next: contact with the building — a door.

[3] Kim, Jo, under revision. [4] Shan, Englot, IROS 2018. [5] Bai et al., IEEE RA-L 2022. [6] Milioto et al., IROS 2019.

2 A planner meets the real world

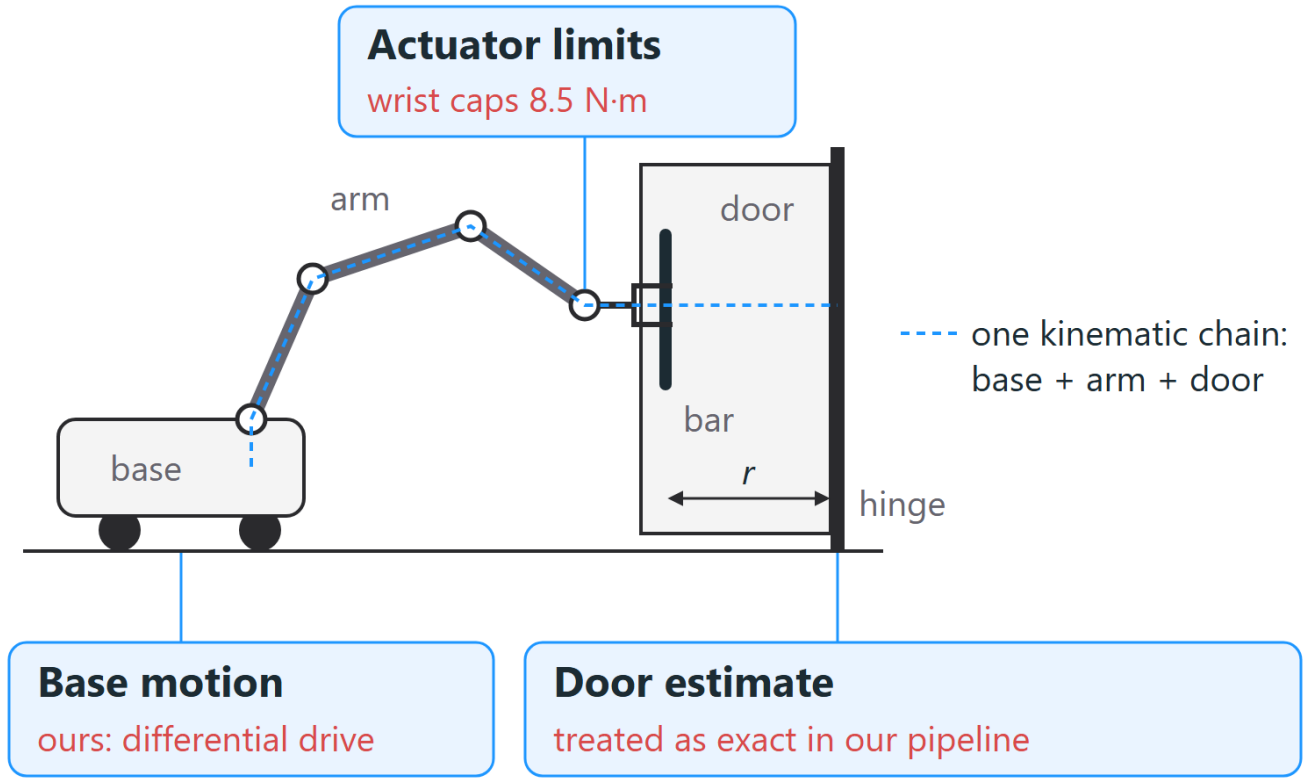
Three gaps in transferring a planner to hardware

AI Robotics Lab

2. A planner meets the real world

One published planner, one real robot

"An augmented kinematic chain meets three hardware gaps"



Base motion

The planner [7] assumes an omnidirectional base. Ours is differential drive.

Door geometry

The pipeline treats the perceived hinge and radius as exact. On glass, they are uncertain.

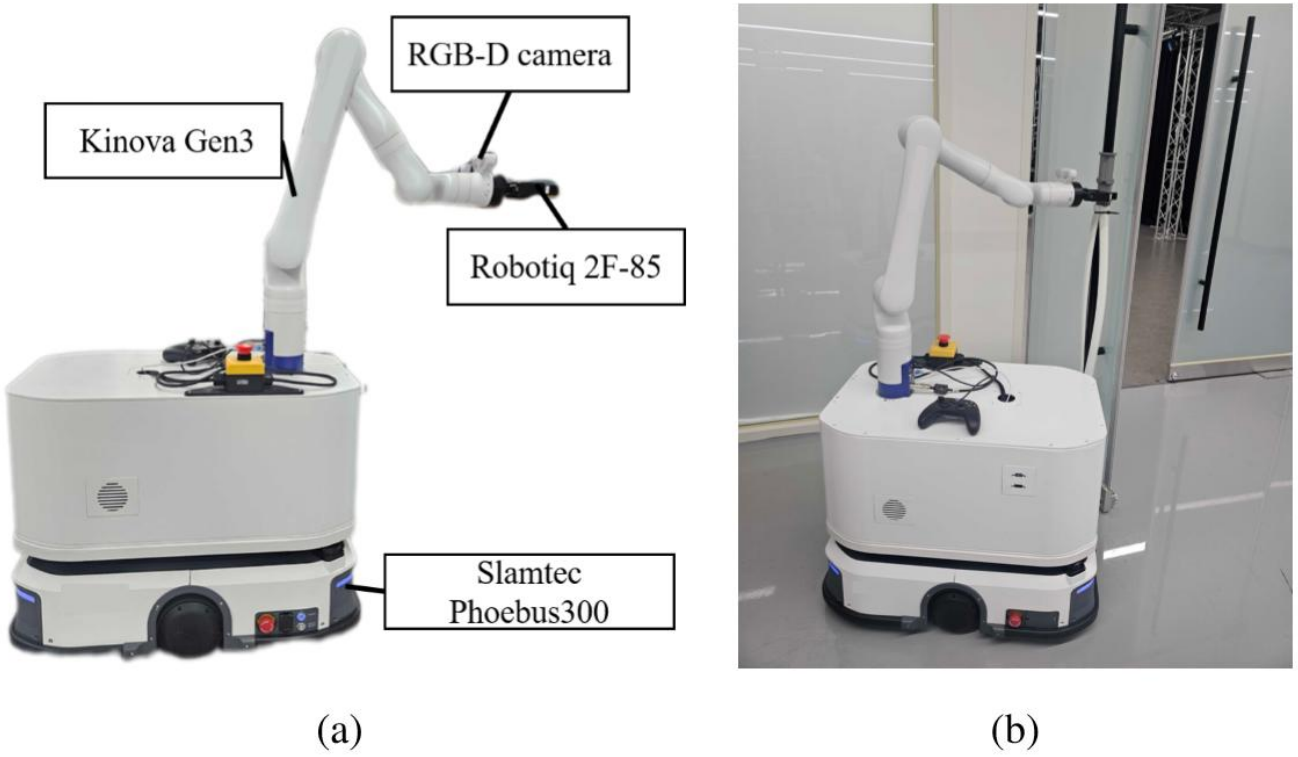
Actuator loading

Actuator load is not modelled. Wrist caps: 8.5 N·m.

[7] Z. Jiao et al., "Integration of robot and scene kinematics for sequential mobile manipulation planning," IEEE Trans. Robot., vol. 41, 2025.

2. A planner meets the real world

The platform [8]



- Slamtec Phoebus300 differential-drive base
- Kinova Gen3 6-DoF arm, Robotiq 2F-85 gripper
- Wrist RGB-D camera, planar lidar
- Rigid sleeve on a 32 mm pull bar
- Guard caps 33 / 8.5 N·m; ~0.2 s over a cap stops the segment

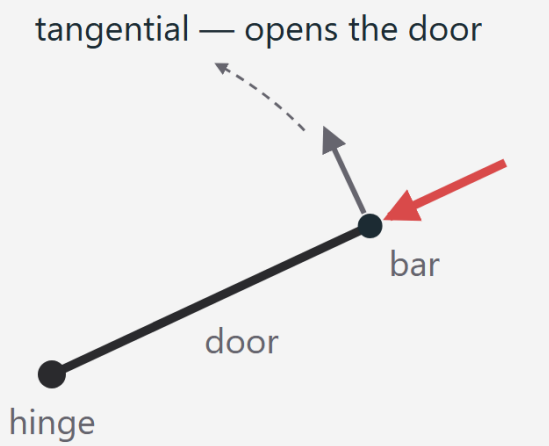
[8] Our manuscript on start-configuration force bounds (ongoing work): Fig. 5.

2. A planner meets the real world

Three preliminary observations on hardware

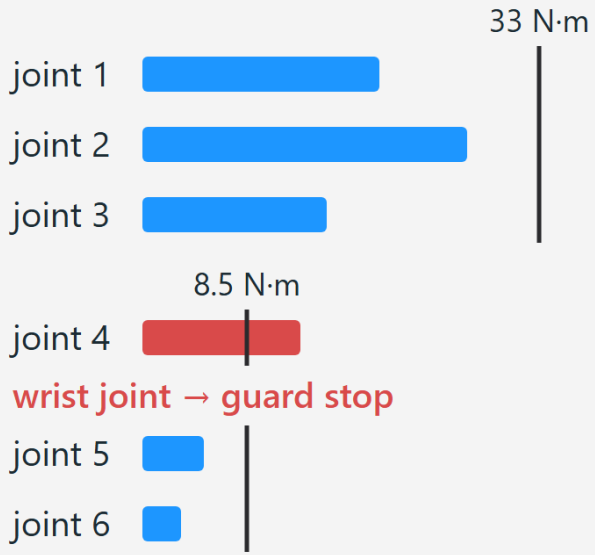
"Kinematic feasibility is not actuation feasibility"

Load at the bar (top view)



radial — larger at each stop

Joint torque vs cap



Initialization

The goal-IK fallback did not converge; a relaxed initialization did (one matched pair).

Door model

Six fitted radii, 613–674 mm; tape 710 mm. The planner takes one as exact.

Actuator load

6 / 6 pulls stopped at the wrist guard. No single hand-force threshold predicts a stop; posture matters.

Preliminary, outside the paper.

3 Torque capability: a ceiling set at the start

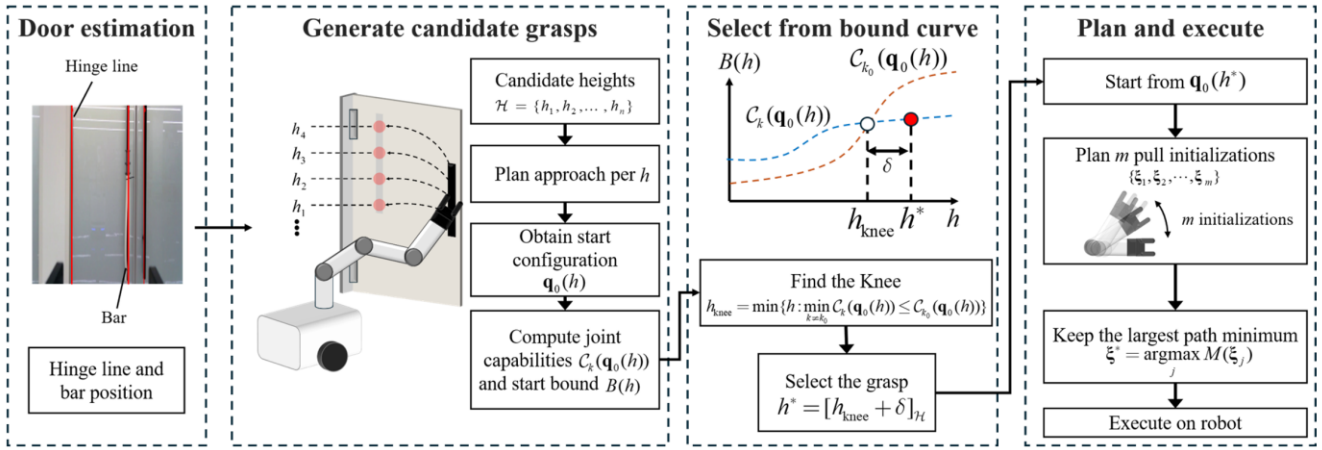
A bound on path minimum capability — ongoing work [8]

AI Robotics Lab

3. Torque limits

What does the start already constrain?

"Grasp and approach fix the first state of every pull"



Pipeline of [8]: estimate door → grasp candidates → select from the bound curve → plan → execute.

Grasp and approach are fixed before trajectory optimization.

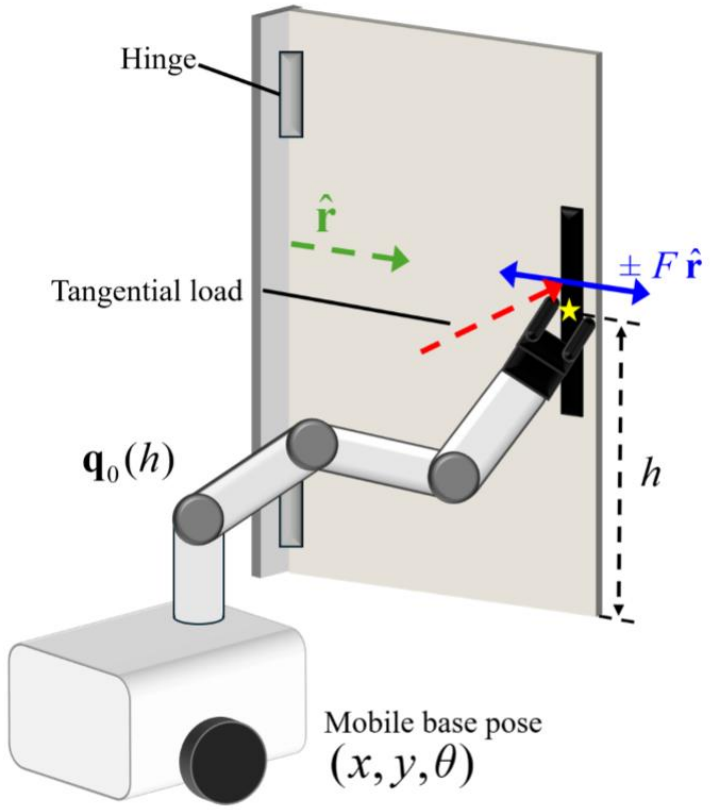
What ceiling does that fixed start impose?

[8] Our manuscript (ongoing work): Fig. 1.

3. Torque limits

Directional force capability at the start

"Capability = remaining torque divided by the lever of a unit radial force"



(4)

$$\tau_k(F) = \tau_k^0 + \ell_k F, \qquad \boldsymbol{\ell}(\mathbf{q}) = \mathbf{J}(\mathbf{q})^\top [\hat{\mathbf{r}};\, \mathbf{0}]$$

(5)

$$\mathcal{C}_k = \frac{\bar{\tau}_k - |\tau_k^0|}{|\ell_k|}, \qquad \mathcal{C} = \min_k \mathcal{C}_k$$

r-hat: horizontal hinge-to-grasp direction; both load signs ±r-hat.

Valid for nonzero lever at a statically feasible posture.

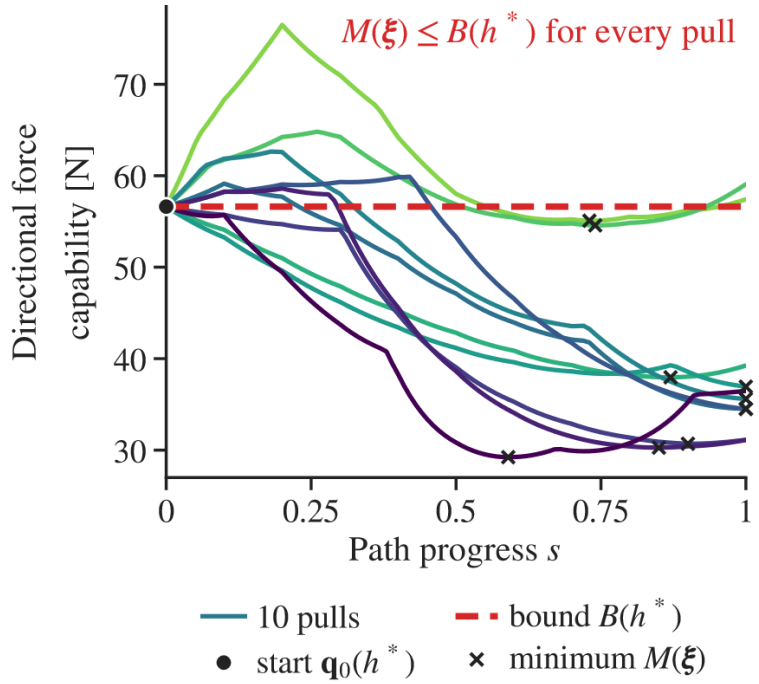
Static torque gravity-only. Caps (33, 33, 33, 8.5, 8.5, 8.5) N·m.

[8] Fig. 2, Eqs. (4)–(5).

3. Torque limits

The start bound: every pull's minimum is at or below it

"Trivial inequality — but it is computed before any pull is solved"



Ten nominal pulls from one start: six exceed B mid-path; no minimum does.

Same capability model for start and path; the pull objective cannot change this.

If F_req > B is required throughout, no pull from this start can satisfy it.

$$B(h) = \mathcal{C}(\mathbf{q}_0(h)), \quad M(\boldsymbol{\xi}) = \min_s \mathcal{C}(\mathbf{q}(s)), \quad M(\boldsymbol{\xi}) \le B(h)$$

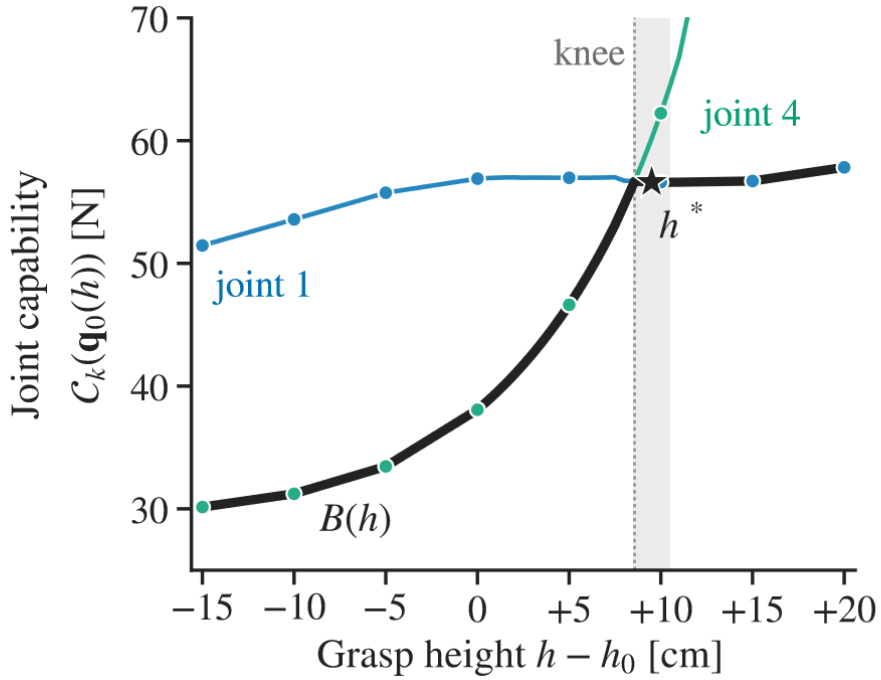
Prop. 1

[8] Fig. 3, Proposition 1.

3. Torque limits

Bound over grasp height: a rise, a knee, a plateau

"Select at the knee plus the placement tolerance, not at the plateau maximum"



Joint 4 limits the rise; joint 1 the plateau. The transition is the knee.

Plateau varies ≤ 1.22 N up to +20 cm.

Knee by interpolation, rounded to the 5 mm grid; offset δ = 10 mm.

Rule and offset fixed before hardware tests.

$$h_{\text{knee}} = \min\{h : \min_{k \neq k_0} \mathcal{C}_k(\mathbf{q}_0(h)) \leq \mathcal{C}_{k_0}(\mathbf{q}_0(h))\}, \qquad h^* = [h_{\text{knee}} + \delta]_{\mathcal{H}} \tag{6}$$

[8] Fig. 4, Eq. (6).

3. Torque limits

Selection before pull optimization

"Compute B with one approach solve per height; solve pulls after selecting a grasp"

			Nominal			RM		
Estimate	Grasp	B	n	E_1	E_5	n	E_1	E_5
A	h_0	38.40	14	32.91	36.84	7	35.32	38.40
	$+5$ cm	47.04	11	40.06	46.50	9	41.72	47.04
	h^*	56.43	27	38.04	50.51	22	44.99	56.02
B	h_0	38.53	9	32.28	36.45	7	36.50	38.53
	$+5$ cm	47.16	12	38.09	45.84	6	41.10	47.16
	h^*	56.47	10	39.30	53.45	7	47.06	56.47
C	h_0	38.08	12	34.16	37.84	9	34.49	38.05
	$+5$ cm	46.61	8	39.42	46.51	9	40.18	46.61
	h^*	56.63	25	35.92	46.98	15	40.72	55.04

+17.9 to +18.5 N

Start-bound gain, gravity-only (47–49 %)

+9.5 cm

Selected height, all three estimates

Approach solve ≈ 0.03 s; pull solve 0.4–1.8 s (per solve).

E5: expected best path minimum among five accepted plans.

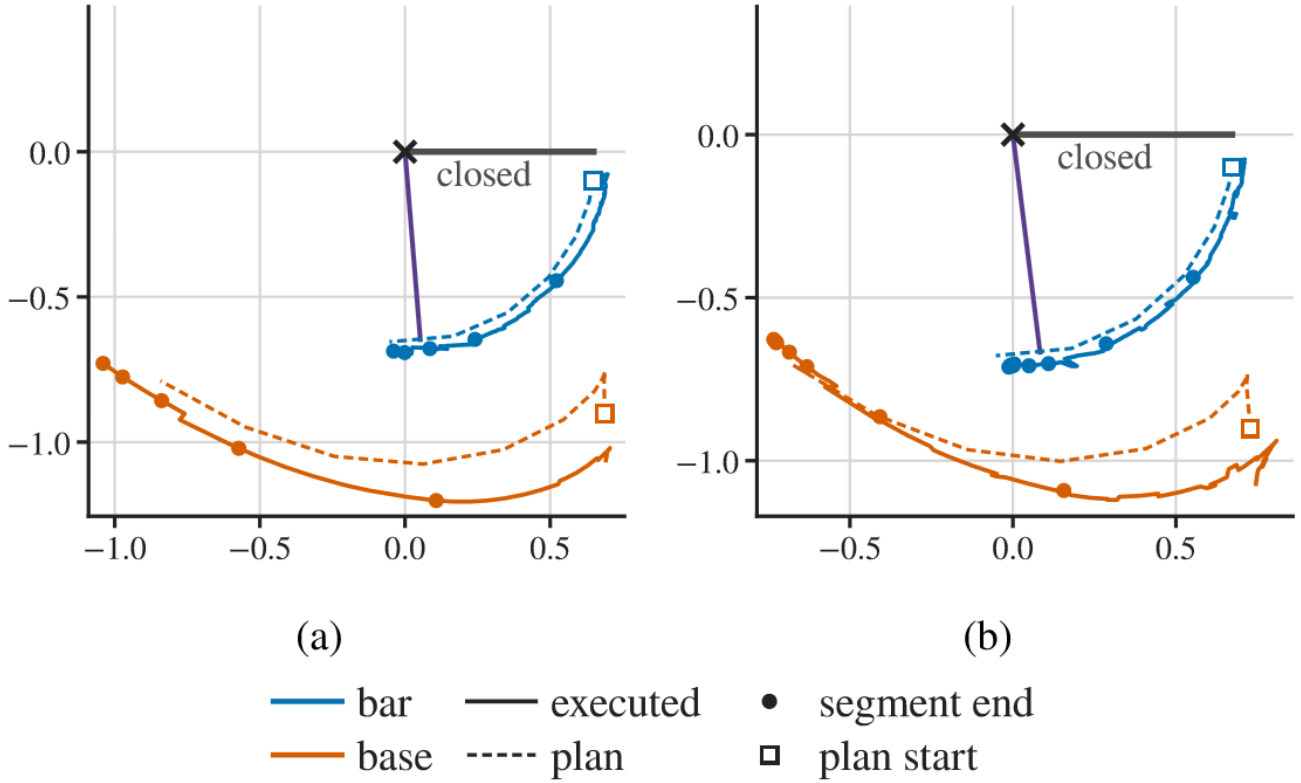
All 12 adjacent comparisons met the corollary condition.

[8] Table I.

3. Torque limits

Hardware 1 · The stance you actually reach

"The bound must be evaluated at the executed stance"



14–18 cm

Base farther from the door than planned (26 pulls)

+0.2 N

Bound gain from height (planned +18.4 N)

A procedure difference, not tracking error.

Joint 1's lever grew to 0.71–0.76 m, so it limited every start at 43–46 N.

Joint 4 still gained: 46–50 → 76–85 N.

[8] Fig. 7.

3. Torque limits

Hardware 2 · Two fixes, then six pulls from the planned stance

"The revised stance reproduced the start-bound gain, with two pulls per height"

			Plan		Execution			
Pull	Obj.	Abs.	B	M	B	M	R	Ended by
Planned stance: 27–28 h_0, 29–30 +5 cm, 31–32 h^*								
27	RM	✓	37.5	37.5	38.0	38.0	14.9	heading
28	RM	✓	37.9	37.9	37.9	37.9	14.8	loop end
29	RM	✓	46.2	46.2	46.7	46.7	29.4	heading
30	RM	✓	46.3	46.3	46.8	46.8	28.9	loop end
31	RM	✓	57.3	57.3	56.1	56.1	23.6	loop end
32	RM	✓	57.7	57.7	56.1	56.1	40.2	heading

9–17 mm
Base start vs plan

0
Joint-guard stops in six pulls

Initial stance only: joint-1 guard stops 9 / 12 without → 1 / 14 with the absorber (observational).

Six RM pulls, two per height. Bounds within 1.6 N of plan; h* raised them by 18.1–18.2 N.

No guard fired. Supports the modelled capability change, not a completion rate.

[8] Table II.

3. Torque limits

Hardware 2 · One execution from the planned stance (video)

"Sweep, approach, grasp, pull — what the numbers look like on the robot"



[8] Execution video. Sweep, approach and grasp at ×32; pull at ×8. Panel: capability C along the pull, dashed line = start bound B.

4 Perception: what the door taught us

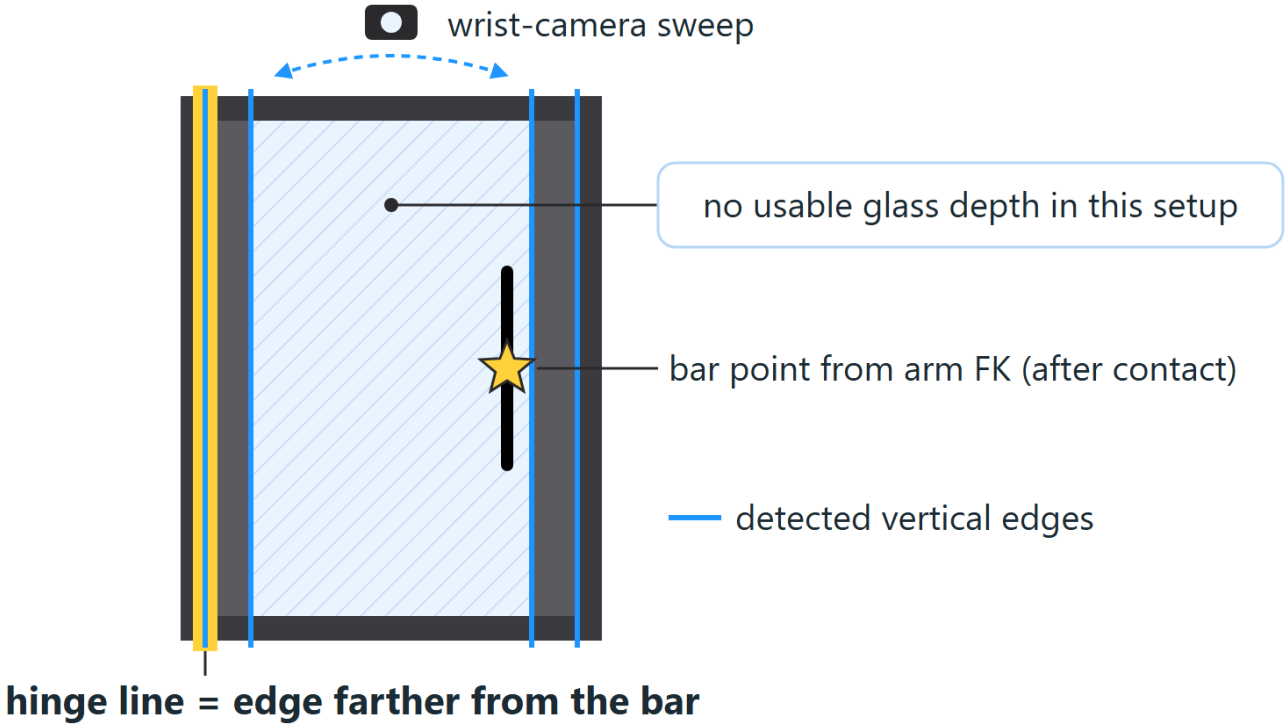
The bound needs one input we never had exactly: the door

AI Robotics Lab

4. Perception

On this glass door, depth was insufficient

"Use the vertical edges of the opaque stiles; the edge farther from the bar is the hinge"



What fails

Glass gave no usable depth.

What we use

Wrist-camera sweep; vertical edges of stiles and frame. The far edge is the hinge.

Bar position

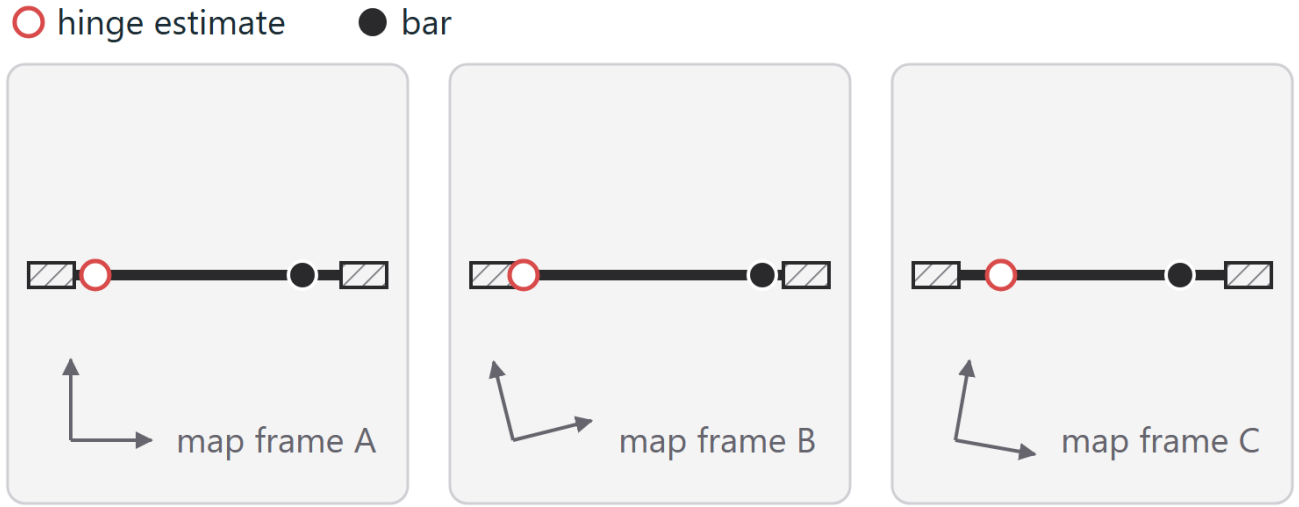
Arm forward kinematics after contact; treated as exact.

[8] Door estimation procedure.

4. Perception

A usable hinge estimate still carries uncertainty

"Hardware ran on a point estimate of the hinge, with no online re-estimation"



between-session coordinate differences up to 337 mm — frames not registered here

point estimates were used; geometric disagreement remains unresolved

28–118 mm
Preliminary hinge deviations; reference unverified

up to 337 mm
Between-session difference, incl. map frames

Point estimate, no online re-estimation. Success here is not a tolerance guarantee.

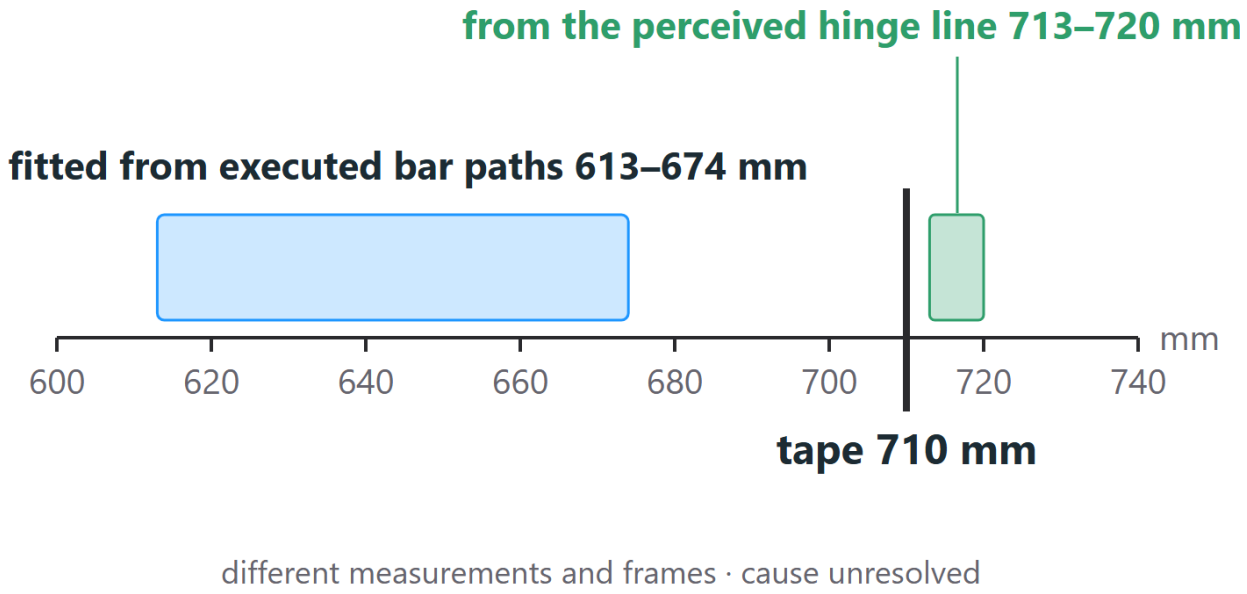
337 mm includes map-frame differences; 28–118 mm are preliminary within-session deviations.

[8] Between-session estimate differences.

4. Perception

Estimated radii and executed-arc fits disagree

"The disagreement is observed; its cause remains unresolved"



different measurements and frames · cause unresolved

Preliminary values outside the paper.

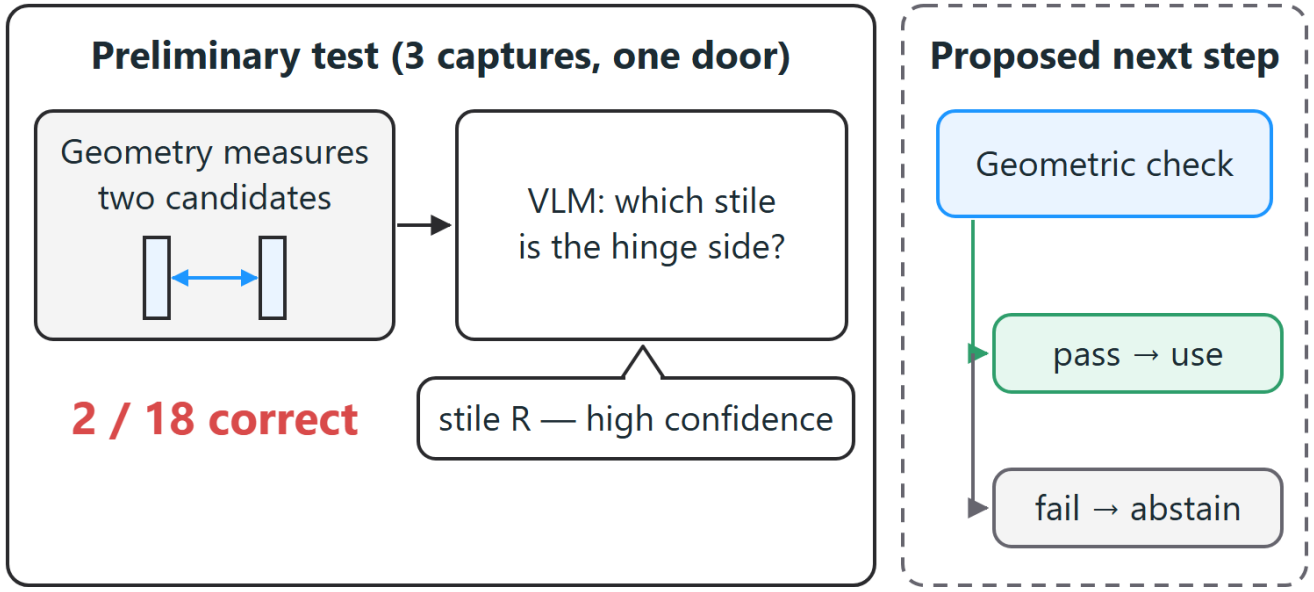
Hinge-derived radii match the tape; executed-arc fits are shorter.

Drift, sleeve motion, frame effects: candidate causes, not isolated.

4. Perception

A VLM geometry judge failed in a small preliminary test

"For the proposed pipeline, verify or abstain before acting on a candidate"



proposed: VLM proposes → geometric check or abstention

2 / 18
Correct, decision-relevant subset

3 captures
One door — a characterization, not a benchmark

Geometry measures candidates; the model picks the semantically right one.

Some wrong answers came with high stated confidence and unanimous votes.

Proposed: the model proposes, geometry judges, else abstain.

5 Exploration and language: a layered world model

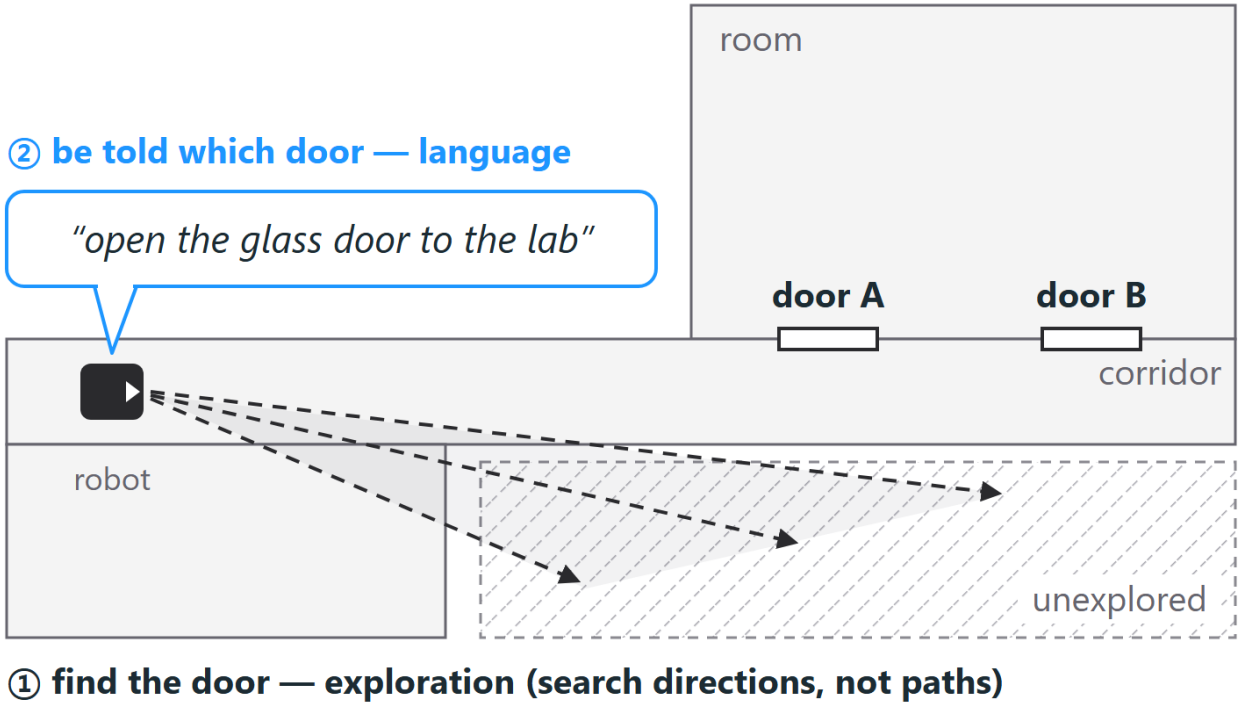
Proposed integration, informed by preliminary offline observations

AI Robotics Lab

5. Exploration and language

Two remaining problems — design

"Two remaining problems: find the door, and be told which door"



① find the door — exploration (search directions, not paths)

Finding the door · exploration
Reach a stance with visibility, reach and modelled torque capability. The bound is a screen, not a guarantee.

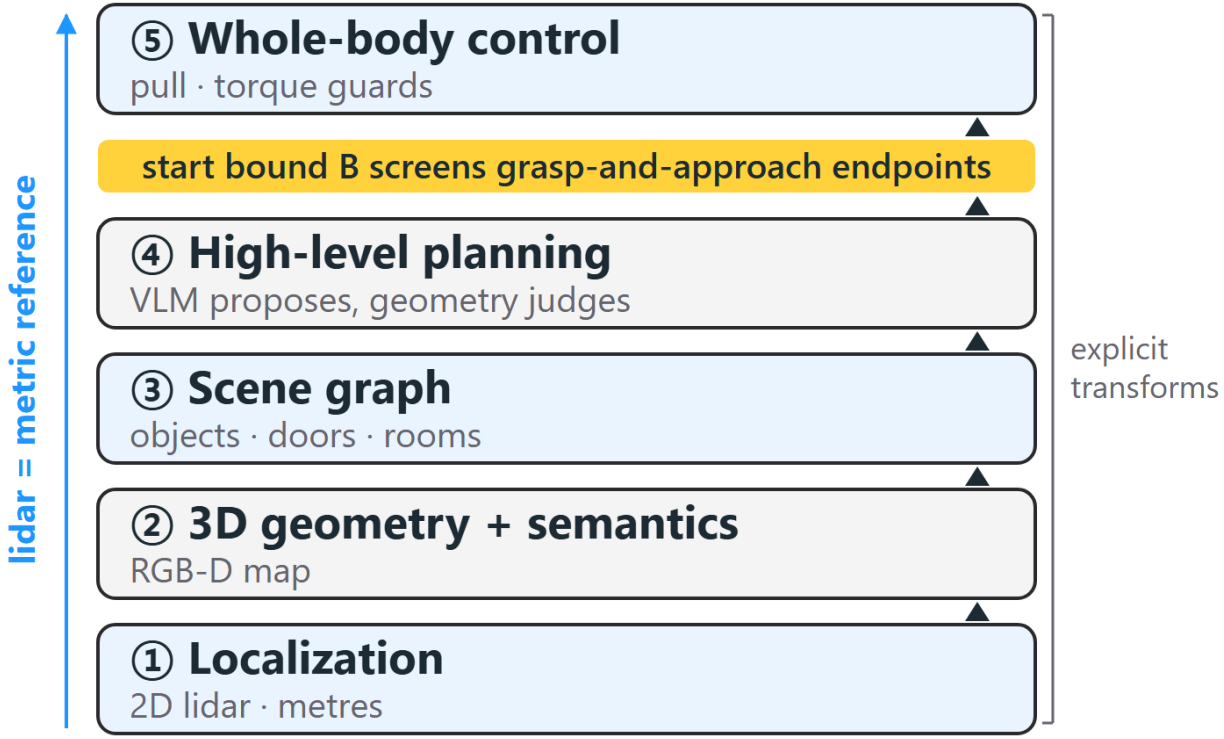
Being told which door · language
A spoken goal → candidates → a judge.

Models propose; geometry and torque checks screen; a human supervises contact.

5. Exploration and language

A layered world model — design

"Use lidar as a metric reference, with vision supplying complementary geometry"



Localization → geometry + semantics → scene graph → planning → control.

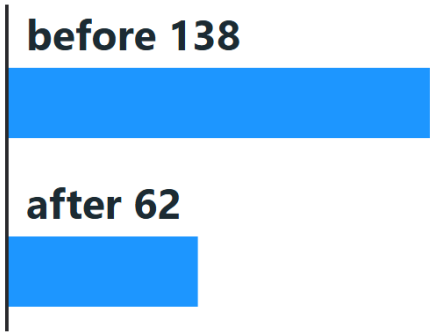
Explicit frames and scale. The bound screens grasp-and-approach endpoints; it is not a frontier score yet.

5. Exploration and language

Frontier selection on recorded data — preliminary replay

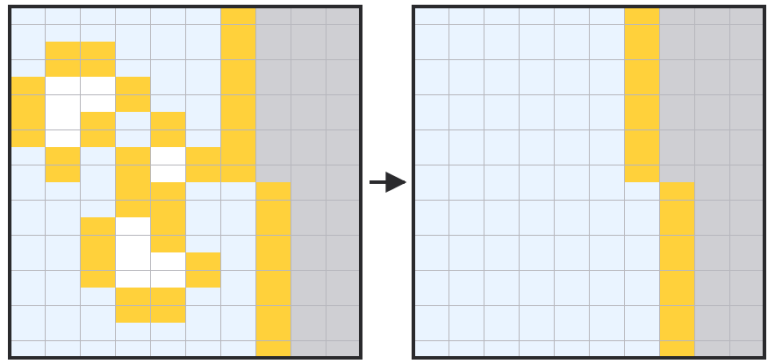
"Depth completion changed frontier counts in one replay setting"

Reported counts in one replay



81 frames ·
depth completion off / on

Possible mechanism



depth holes may create extra frontiers

after filling holes

frontier cell · depth hole · unexplored · known free

replay on recorded data · loop not closed on the robot

138 → 62

Frontiers before / after depth completion (81 frames)

same target

Target unchanged across tested weights

OpenFrontier [9] on our recorded data.

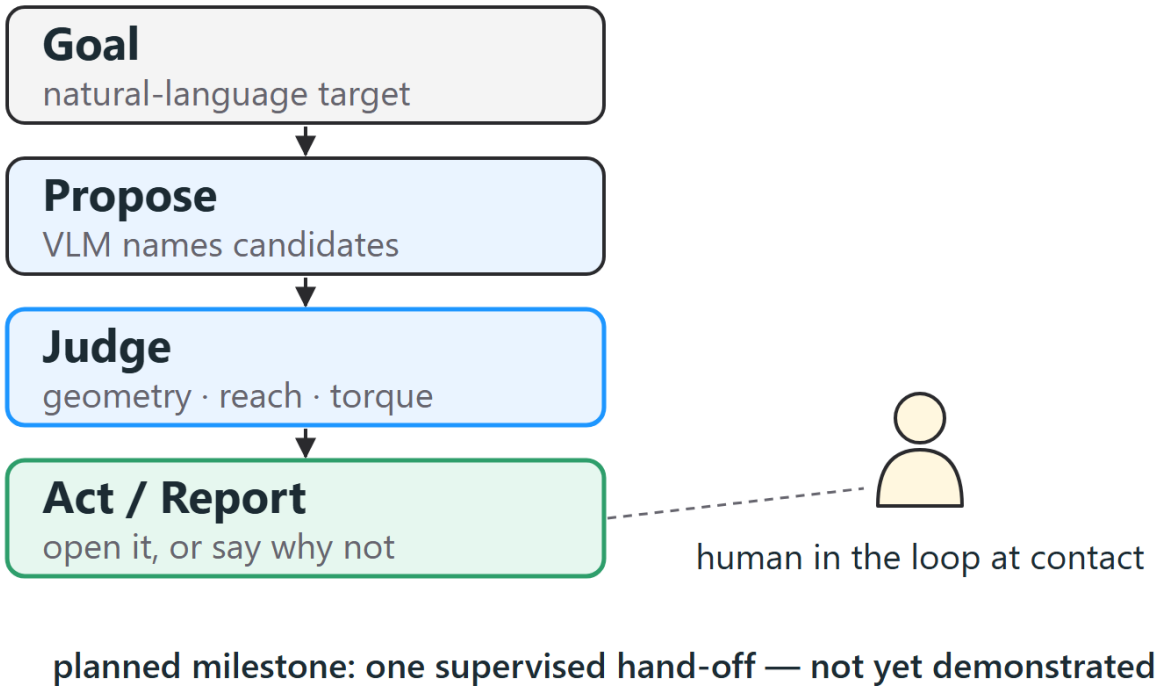
Consistent with depth holes counted as frontiers; cause and frontier quality not established.

[9] E. Padilla-Cerdio*, B. Sun*, M. Pollefeys, H. Blum, "OpenFrontier: General Navigation with Visual-Language Grounded Frontiers," RSS 2026 (arXiv:2603.05377).

5. Exploration and language

Where language enters — plan

"First-year deliverable: one supervised hand-off demo — explore, find the door, open it"



planned milestone: one supervised hand-off — not yet demonstrated

Milestone: one end-to-end handoff with human supervision at contact. Not yet demonstrated.

6 Closing

What was measured, and what is next

AI Robotics Lab

6. Closing

What was measured

"Select by the start-capability bound, then evaluate it at the executed stance"

Hardware transfer

Three planner assumptions became three measured gaps: initialization, door geometry, actuator load.

Start capability

Selected height raised B by ~18 N offline and in six pulls. A capability gain, not a success rate.

Perception and next steps

Point estimates stay uncertain; VLM checks were unreliable; the language pipeline is proposed.

6. Closing

What is next

Evaluate the capability objective

Test direct optimization of path minimum capability.

Test the integrated handoff

Screen sensing, reach and torque; connect exploration to supervised door pulling.

Other articulated objects: a preliminary drawer test suggests transfer; it needs its own evaluation.

Thank you

AI Robotics Lab

KangGeon Kim
AI Robotics Lab, Jeonbuk National University