# Locomotion in Human Spaces: Perception-Aware Whole-Body Stabilization for HRI

Md Hafizur Rahman
Control & Instrumentation Department, King Fahd University of Petroleum & Minerals
Dhahran, Saudi Arabia
g202319450@kfupm.edu.sa

Tansu Sila Haque
Control & Instrumentation Department, King Fahd University of Petroleum & Minerals
Dhahran, Saudi Arabia
g202501950@kfupm.edu.sa

Muhammad Faizan Mysorewala
Control & Instrumentation Department, King Fahd University of Petroleum & Minerals
Dhahran, Saudi Arabia
mysorewala@kfupm.edu.sa

## Abstract

Mobile interaction at arm's length demands motions that are precise, comfortable, and readily legible to people. Yet many pipelines decouple navigation from human-facing alignment, leaving last-meter oscillations and stance jitter unaddressed. This work introduces a quadruped *center–stop* (CS) pipeline that integrates perception and control for close-range approach: (i) a real-time detector with depth association keeps a human-held object near image center; (ii) a *freshness gate* rejects stale RGB–D estimates, preventing twitch from intermittent frames; (iii) a bounded *safety bubble*—a stop band around the desired range and image center—freezes base motion and enforces a brief dwell; and (iv) a velocity-level whole-body balance module filters desired base twists via a small QP, keeping a CoM proxy within a shrunken support polygon while applying IMU-based yaw/tilt damping so the stance appears quiet and predictable. Hardware validation quantifies centering accuracy, settle time, jerk-based smoothness, and perceived comfort/legibility; complementary simulation probes robustness to injected pose dropouts up to 300 ms under matched camera and controller limits. Results show consistent near-center composition, rapid convergence into the safety bubble without limit cycles, reduced near-target jerk, and stable holding under brief perception gaps. The CS detection-to-approach pipeline and explicit safety bubble provide a practical, reproducible template for socially acceptable, close-range navigation around people on legged bases.

## CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Computer systems organization** → **Robotics**; • **Theory of computation** → *Control theory.*

## Keywords

human–robot interaction, safety bubble, quadruped locomotion, whole-body control, RGB–D perception, freshness gating


ACM ISBN 978-1-4503-XXXX-X/26/03
https://doi.org/10.1145/XXXXXXX.XXXXXXX



**ACM Reference Format:**
Md Hafizur Rahman, Tansu Sila Haque, and Muhammad Faizan Mysorewala. 2026. Locomotion in Human Spaces: Perception-Aware Whole-Body Stabilization for HRI. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '26), March, 2026, Scotland, UK*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/XXXXXXX.XXXXXXX

## 1 Introduction

Humans infer intent from motion almost instantly. As a robot closes the last meter to a person, small differences in timing, lateral alignment, and stance stability dominate perceived safety and competence. On legged platforms these aspects are tightly coupled: base stance changes are visually salient; modest perception delays can excite oscillations; and any twitch near the body is magnified in peripheral vision. Despite rapid progress in legged locomotion, close-range interaction remains underexplored in HRI because navigation and manipulation are often engineered as separate subsystems with different objectives and bandwidths. The result is last-moment corrections and stance readjustments that undermine comfort and legibility [9, 11, 23, 38].

This work studies a common exchange: a quadruped approaches a person who is holding a bottle and *stops* at a nominal, socially acceptable standoff $d^*$. The desired behavior is calm and easily interpretable on first exposure, even with intermittent detections. The approach is platform-agnostic and relies only on image-plane centering, a depth-derived range estimate, and a *velocity-level whole-body balance module* that filters base motion for a visually steady stance. Experiments are conducted on a Unitree GO2 with a head-mounted RGB–D camera (the onboard 7-DoF arm is present but unused in this study, see the Figure 1).

Three principles guide the controller design:

- **Legibility over optimality.** Motions should make intent obvious rather than merely kinematically efficient. The controller biases toward smooth, minimum-jerk-like changes and a consistent visual framing of the handheld object [9, 11].
- **Posture steadiness at interaction range.** Within arm's length, base oscillations translate directly into perceived risk. We therefore include a lightweight Whole-Body *balance* module that (i) keeps a proxy of the center of mass within a shrunken support polygon, (ii) applies IMU-based yaw damping and tilt gating, and (iii) respects near-human speed limits. The module solves a small convex QP with OSQP when available [21, 36] and otherwise falls back to a projection-onto-convex-sets routine; unlike full dynamics WBC [18, 31, 34], it operates at the twist (velocity) level to stabilize stance during approach and hold.

**Figure 1: Unitree Go2 Robot with an 7-DoF arm.**

- **Perception-aware control.** RGB–D estimates exhibit nontrivial age and occasional dropouts. A *freshness gate* rejects stale poses and holds the last valid target to prevent micro-corrections that humans notice near the body.

Behavior is organized as a simple, perception-aware *center–stop* policy (pipeline in Fig. 2, FSM in Fig. 3). First, the robot *centers* the human-held object using a bounded image-based controller that respects near-human comfort limits on forward/lateral speeds and yaw rate. Second, when the object is near the image center and the range is within a tight window around $d^*$, the robot *stops* base motion and dwells inside a rectangular *stop band*. This deliberate inhibition creates a safety bubble and removes last-meter limit cycles introduced by discretization and perception latency. A recovery state freezes motion if estimates go stale, then resumes approach once fresh data return. The whole-body balance module runs continuously underneath to filter the commanded twist so that the predicted CoM remains inside the support polygon and yaw/tilt transients are attenuated—further reducing near-person twitch.

Evaluation focuses on the most safety-critical region for HRI—closing to within arm's length. Controlled hardware trials quantify centering accuracy, smoothness (jerk proxies), and stop-band stability, together with subjective ratings of comfort and legibility. Stabilizer-specific metrics (polygon slack/violations, yaw attenuation, residual motion during hold) isolate the contribution of the balance module relative to an unfilted baseline. Ablations test the role of the freshness gate, stop-band dwell, and stabilizer options (projection vs. QP), linking each component to HRI-facing outcomes.

*Contributions.*

(1) **HRI–centric *center–stop* policy.** A perception–aware control scheme that *operationalizes proxemics* via an explicit safety bubble: a rectangular stop band in image–range space with a short dwell and near-human speed bounds, yielding approach motions that are comfortable, predictable, and legible at arm's length.

(2) **Temporal conditioning for human-perceived stability.** A lightweight *freshness gate* (with optional EMA smoothing) that admits only age-bounded detections and holds targets under dropouts, suppressing micro-corrections and last-meter jitter that observers find unsettling.

(3) **HRI-aligned evaluation protocol.** Objective metrics tied to human factors—centering accuracy, stop-band entry/dwell, residual motion, and jerk proxies—paired with subjective ratings of comfort, perceived safety, and legibility; ablations isolate the effects of the safety bubble and temporal gating.

(4) **Velocity-level whole-body balance module.** A small-footprint QP filter (Sec. 3.3) that maps desired base twists to stabilized commands by keeping a CoM proxy within a shrunken support polygon, enforcing speed bounds, and applying IMU-based yaw/tilt damping; this yields a visually steady stance during StopBandHold and enhances perceived stability.

(5) **Reproducibility and safety disclosure.** Complete controller equations, finite-state logic, TF/ROS 2 interfaces, and parameter tables enabling replication on legged bases, with explicit speed/distance safeguards for human-space operation.

## 2 Related Work

Close-range interaction between a mobile manipulator and a person sits at the intersection of social navigation, perception under latency, and control of whole-body posture at arm's length. While each strand is well studied, comparatively few systems bind them end-to-end for legged platforms that must *approach and stop* near people with motions that look legible and feel comfortable.

Human observers infer intent from the geometry and timing of robot motion; legibility and predictability have been formalized and evaluated in controlled HRI studies [2, 9]. A long tradition also links human preference to minimum-jerk and related smoothness models [11]. These ideas motivate velocity bounds and visual-centering strategies that signal intent during approach (Sec. 3.2).

Proxemic conventions and social distances for approach are central to comfort [3, 20, 35, 38]. Crowd-aware navigation models interaction explicitly [37]. Broader HRI safety surveys emphasize limiting kinetic energy, reducing surprises, and designing for subjective comfort [13, 22, 23]. The present work implements these principles with speed limits, a safety bubble (stop band with dwell), and temporal gating (Alg. 1).

Operational-space control [18] and its descendants coordinate posture and contacts; representative whole-body / QP formulations

**Table 1: Representative strands versus the present *center–stop* strategy.**

| Area | Representative works | Typical gap at arm's length / This paper |
|---|---|---|
| Legible & smooth motion | [2, 9, 11] | Often no explicit mechanism to prevent last-meter oscillations. *Here:* bounded visual centering plus an inhibition band with dwell. |
| Proxemics & social navigation | [3, 20, 35, 37, 38] | Focus on path-level distances, not the final decimeter. *Here:* a safety bubble ((6)) enforced online. |
| HRI safety (surveys/standards) | [1, 13, 22, 23] | Guidance on limits, less on perception-latency effects. *Here:* freshness gate (Alg. 1) to remove stale-driven twitch. |
| Operational-space & WBC | [8, 12, 18, 21, 25, 26, 31, 34] | Provide posture/contact coordination. *Here:* used purely as a stabilizer during approach/hold (Sec. 3.3). |
| Perception for hand-held objects | [5, 17, 30, 39] | Real-time but flickery at close range. *Here:* inner-crop median depth and temporal conditioning (Sec. 3.1). |
| Quadruped locomotion & loco-manipulation | [4, 6, 12, 16, 29, 32, 33] | Mobility is strong; close-range HRI is under-reported. *Here:* explicit last-decimeter policy for legged bases. |

include [8, 21, 25, 26, 31, 34]. For legged systems, whole-body stabilization during stance is mature [12]. Our controller uses these conventions as a base stabilizer with a high weight on attitude/height regulation (Sec. 3.3), solved with modern dynamics and QP libraries [7, 36].

ANYmal, HyQ, Cheetah and related families demonstrate robust mobility [4, 6, 16, 33]. Early loco-manipulation on legged robots highlights feasibility but often decouples navigation and close-range interaction [29, 32]. The *center–stop* policy specifically targets the last decimeter near a person, where decoupled designs tend to oscillate.

Single-shot detectors provide real-time boxes and confidences [5, 30, 39]; consumer RGB–D cameras provide synchronized depth [17]. In close-range HRI, intermittent estimates can provoke “micro-corrections.” The freshness gate (Alg. 1) implements a simple temporal condition to prevent stale frames from driving the controller.

The ROS 2 navigation stack offers pragmatic goal-directed autonomy [24]. For optimization, switched-systems and MPC toolkits such as OCS2 are commonly used in legged control [10]. Although the present approach uses a lightweight feedback policy (Sec. 3.2), these tools inform the broader landscape.

Even for approach-only behavior, compliance concepts influence perceived safety. Impedance control—original and passivity-based formulations [14, 15, 27]—and series elasticity [28] motivate conservative velocity bounds and dwell policies consistent with ISO/TS 15066 guidance [1]. Semantic legibility ideas similarly emphasize interpretable motion primitives [19].

*Problem Statement.* Consider a quadruped base with a head–mounted RGB-D camera interacting with a person who holds a bottle. From an initial separation of roughly 2.5 m, the robot must execute a perception–aware *center–stop* policy that:

(1) **Approaches with legibility.** Keep the bottle near the image center while closing range, using the bounded controller in Sec. 3.2 to respect near–human comfort limits on forward/lateral speeds and yaw rate; command normalization makes pixel errors comparable across distance and field of view.
(2) **Stops inside a safety bubble.** Enter and remain within a rectangular inhibition band in image–range space centered at the nominal standoff $d^*$, with half–widths $(\epsilon_x, \epsilon_y, \epsilon_d)$; freeze base motion and dwell for $\tau_{\text{hold}}$ to eliminate last–meter oscillations (Fig. 3).
(3) **Maintains a steady stance.** Hold a visually quiet posture while in the bubble by filtering desired twists through the velocity–level whole–body balance module (Sec. 3.3), which keeps a CoM proxy inside a shrunken support polygon and applies IMU–based yaw/tilt damping.

Operational constraints include: (i) near–human speed bounds on $(v_x, v_y, \omega_z)$; (ii) robustness to RGB-D dropouts up to 300 ms using the *freshness gate* (Alg. 1) with zero–order hold and optional EMA smoothing; and (iii) a minimum distance $d_{\text{min}}$ that prevents intrusion into personal space while approaching $d^*$. The task *excludes manipulation*; evaluation therefore centers on approach accuracy, jerk–based smoothness, stop–band stability (entry, dwell, residual motion), and subjective ratings of safety, comfort, and legibility (Sec. 4).

# 3 Method and Implementation

This section formalizes the *center–stop* strategy implemented in the released ROS 2 nodes for close-range interaction with a person who is holding a bottle. The focus is HRI: motions should be legible, speed-bounded, and free of last-meter oscillations. Figures 2 and 3 visualize the dataflow and the phase logic; Algorithm 1 summarizes the temporal gate that prevents stale-perception jitter. All symbols are defined where they first appear, and default values appear in Table 2.

## 3.1 Perception: Bottle-in-Hand Detection, Depth, Pose, and Temporal Conditioning

*2D detection and bottle–human association.* A real-time YOLO-style model returns axis-aligned boxes $\mathcal{B} = [u_{\min}, u_{\max}, v_{\min}, v_{\max}]$ with class label and confidence $\gamma$ [5, 30, 39]. The 2D target pixel is the box centroid

$$(u, v) = \left(\tfrac{u_{\min}+u_{\max}}{2},\ \tfrac{v_{\min}+v_{\max}}{2}\right),$$

and low-confidence detections with $\gamma < \gamma_{\min}$ are rejected. To keep the task HRI-relevant, a `bottle` is accepted only if it overlaps a `person` box or lies within a proximity band $\delta_{\text{hand}}$ of the person's lower third (a proxy for the hand/forearm region). Among valid candidates, the detector chooses the one that maximizes

$$c \;=\; \gamma \;-\; \lambda_{\text{edge}}\,\frac{\|(u,v) - (c_x, c_y)\|_2}{\sqrt{W^2 + H^2}},$$

which softly prefers central framing, where $(c_x, c_y)$ is the image center and $W \times H$ is the image size.

*Depth association and robust range.* Let $Z(u, v)$ denote the RGB–D depth registered to the color stream. A robust range estimate is the median over an inner crop $\mathcal{B}_\eta$ (shrunken box, $\eta \in [0, 1]$):

$$d \;=\; \operatorname{median}\Big\{ Z(u_i, v_i) \;\Big|\; (u_i, v_i) \in \mathcal{B}_\eta,\; Z(u_i, v_i) \in [d_{\min}, d_{\max}] \Big\}. \tag{1}$$

Only finite depths within $[d_{\min}, d_{\max}]$ (e.g., $[0.2, 4.0]$ m) are used to suppress edge/hand contamination and shiny-object outliers.

*Back-projection and target in* `base_link`. With intrinsics $K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$, the 3D point in the camera frame is

$$\boldsymbol{p}_c \;=\; d\, K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \tag{2}$$

and in the base frame $\{b\}$ it is

$$\boldsymbol{p}_b \;=\; {}^{b}R_c\, \boldsymbol{p}_c + {}^{b}\boldsymbol{t}_c, \tag{3}$$

where ${}^{b}R_c \in \mathbb{R}^{3\times 3}$ and ${}^{b}\boldsymbol{t}_c \in \mathbb{R}^3$ are calibrated camera→base extrinsics. Two scalar image-plane errors and a range error are then defined as

$$e_x = u - c_x, \quad e_y = v - c_y, \quad e_d = d - d^*,$$

with $d^*$ the nominal standoff distance (the center of the safety bubble).

*Temporal conditioning (freshness and smoothing).* RGB–D pipelines can be delayed or flickery. Two lightweight devices stabilize the stream with minimal latency:

**Freshness gate.** A pose $(\hat{\boldsymbol{p}}, t)$ is *fresh* if its age $\Delta t = t_{\text{now}} - t$ satisfies $\Delta t \le T_{\text{fresh}}$; otherwise the controller *holds* the last valid target (zero-order hold). This avoids human-salient "micro-corrections." See Algorithm 1.

**EWMA smoothing.** Accepted poses are blended with an exponential moving average (EMA) to reduce pixel jitter without delaying phase decisions:

$$\boldsymbol{p}_b \leftarrow (1-\alpha)\, \boldsymbol{p}_b^{\text{prev}} + \alpha\, \hat{\boldsymbol{p}}_b, \qquad \alpha = 1 - e^{-\Delta t/\tau}. \tag{4}$$

Typical values are $T_{\text{fresh}} = 300$ ms and $\tau = 0.12$ s.

**Algorithm 1** Freshness gate with zero-order hold and optional EMA

```
Parameters: T_fresh (age window), τ (EMA time constant)
State: target ∈ R^3 (last valid in {b}) or None
while node running do
    (p̂_b, t, γ) ← Detect&BackProject()
    if γ ≥ γ_min and t and t_now − t ≤ T_fresh then
        α ← 1 − e^{−(t_now − t)/τ}
        target ← (1 − α)·target + α · p̂_b        ▷ EWMA
    end if
    publish(target)        ▷ Zero-order hold when no fresh update
end while
```

### 3.2 Center–Stop Control (legible approach with a safety bubble)

*Center (alignment with comfort bounds).* Given image errors $(e_x, e_y)$ and range error $e_d$, the base command is

$$\begin{aligned} v_x &= \operatorname{sat}_{v_{x,\max}}(k_x\, e_d), \\ v_y &= \operatorname{sat}_{v_{y,\max}}\!\big(k_y\, \tfrac{d}{f_x}\, e_x\big), \qquad \omega_z = \operatorname{sat}_{\omega_{\max}}\!\big(k_\psi\, \tfrac{1}{f_y}\, e_y\big), \end{aligned} \tag{5}$$

where $v_x$ (forward), $v_y$ (lateral), and $\omega_z$ (yaw rate) are saturated by the comfort limits $(v_{x,\max}, v_{y,\max}, \omega_{\max})$, and $\operatorname{sat}_c(x) = \operatorname{sign}(x)\min(|x|, c)$. The factors $d/f_x$ and $1/f_y$ normalize pixels into roughly metric lateral/yaw corrections, preserving similar visual centering across distance and field of view—improving legibility to the human partner.

*Stop (oscillation suppression inside the personal-space band).* Define a rectangular inhibition band in the image–range space:

$$\mathcal{B} = \Big\{ (e_x, e_y, e_d) : \; |e_x| \le \epsilon_x,\; |e_y| \le \epsilon_y,\; |e_d| \le \epsilon_d \Big\}. \tag{6}$$

When $(e_x, e_y, e_d) \in \mathcal{B}$, the controller *zeros* the base command,

$$(v_x, v_y, \omega_z) = (0, 0, 0),$$

and runs a dwell timer $\tau_{\text{hold}}$ before declaring a stable "hold." With bounded gains in (5), the discrete-time error dynamics are contractive outside $\mathcal{B}$ and invariant inside, eliminating last-meter limit cycles that observers often interpret as hesitation. The center of the band uses the desired standoff $d^*$; a global minimum distance $d_{\min}$ prevents intrusion into personal space even when depth is noisy.

### 3.3 Whole-Body Control (velocity-level stabilizer)

We deploy a *velocity-level, convex whole-body stabilizer* that filters the base twist so a proxy center of mass (CoM) remains inside a shrunken support polygon while respecting near-human speed bounds and inertial cues. The node runs after the Center–Stop policy: it receives $\boldsymbol{v}_{\text{des}} = [v_x,\, v_y,\, \omega_z]^\top$ and publishes a stabilized $\boldsymbol{v}$ on `/cmd_vel_stab`. When OSQP is available it solves a tiny QP in real time; otherwise it falls back to a projection-onto-convex-sets (POCS) routine. This keeps the stance visually steady during StopBandHold without requiring full rigid-body dynamics.

*Support polygon and prediction.* From TF foot poses $\mathcal{P} = \{\boldsymbol{p}_i = [x_i, y_i]^\top\}_{i=1}^N$ (counterclockwise in $\{b\}$), each edge $e_i : \boldsymbol{p}_i \to \boldsymbol{p}_{i+1}$ defines an inward half-space $\mathcal{H}_i = \{\boldsymbol{p} : \boldsymbol{n}_i^\top \boldsymbol{p} \le b_i\}$ with

$$\boldsymbol{n}_i = \frac{1}{\|\bar{\boldsymbol{n}}_i\|_2} \begin{bmatrix} y_{i+1} - y_i \\ -(x_{i+1} - x_i) \end{bmatrix}, \qquad b_i = \boldsymbol{n}_i^\top \boldsymbol{p}_i. \tag{7}$$

A safety margin shrinks the polygon by $\delta_{\text{poly}}$ via $b_i \leftarrow b_i - \delta_{\text{poly}}$. The CoM is proxied by a fixed offset $\boldsymbol{p}_{\text{com}}^0 = [x_{\text{off}},\, y_{\text{off}}]^\top$, and its short-horizon displacement is modeled quasi-statically as

$$\boldsymbol{p}_{\text{com}}^{\text{pred}}(\boldsymbol{v}) = \boldsymbol{p}_{\text{com}}^0 + \underbrace{\begin{bmatrix} \beta_x \Delta t & 0 & 0 \\ 0 & \beta_y \Delta t & 0 \end{bmatrix}}_{\mathbf{B}} \boldsymbol{v}, \tag{8}$$

**Algorithm 2** Velocity-level whole-body stabilizer (QP with POCS fallback)

**Require:** $\boldsymbol{v}_{\text{des}}$, IMU $(\omega_z^{\text{imu}}, \theta_{\text{tilt}})$, foot TFs $\mathcal{P}$, params $\{\delta_{\text{poly}}, \Delta t, \beta_x, \beta_y, \mathbf{W}, w_z, \boldsymbol{v}_{\min}, \boldsymbol{v}_{\max}\}$
Build $\mathbf{A}$, $\boldsymbol{u}$ from $\mathcal{P}$ using (7)–(9)
$\tilde{\boldsymbol{v}}_{\text{des}} \leftarrow$ yaw-damped (10); apply tilt gate (11)
**if** OSQP available **then**
    Solve (14)–(15); set $\boldsymbol{v}^\star$
**else**
    $\boldsymbol{v} \leftarrow \Pi_{[\boldsymbol{v}_{\min}, \boldsymbol{v}_{\max}]}(\tilde{\boldsymbol{v}}_{\text{des}})$; for $K = 2{:}3$ sweeps: project onto $\boldsymbol{a}_i^\top \boldsymbol{v} \le u_i$ and re-clamp; set $\boldsymbol{v}^\star$
**end if**
Publish $\boldsymbol{v}^\star$ on `/cmd_vel_stab`

with gains $\beta_x, \beta_y > 0$. Requiring $\boldsymbol{n}_i^\top \boldsymbol{p}_{\text{com}}^{\text{pred}}(\boldsymbol{v}) \le b_i$ yields the linear inequalities

$$\underbrace{\boldsymbol{n}_i^\top \mathbf{B}}_{\boldsymbol{a}_i^\top} \; \boldsymbol{v} \le \underbrace{b_i - \boldsymbol{n}_i^\top \boldsymbol{p}_{\text{com}}^0}_{u_i}, \qquad \Rightarrow \qquad \mathbf{A}\boldsymbol{v} \le \boldsymbol{u}. \tag{9}$$

*Inertial guards and comfort bounds.* IMU yaw-rate damping tempers spin near people:

$$\tilde{v}_{\text{des}} = \begin{bmatrix} v_x \\ v_y \\ \omega_z - k_d^{\omega}\, \omega_z^{\text{imu}} \end{bmatrix}. \tag{10}$$

Let $\theta_{\text{tilt}} = \sqrt{\phi^2 + \vartheta^2}$ (roll–pitch magnitude). Forward speed is gated against tilt:

$$v_x \leftarrow \begin{cases} 0, & \theta_{\text{tilt}} \ge \theta_{\max}, \\ s(\theta_{\text{tilt}})\, v_x, & \theta_{\text{tilt}} \in (\theta_{\text{soft}}, \theta_{\max}), \\ v_x, & \text{otherwise}, \end{cases} \qquad s(\theta) = 1 - \frac{\theta - \theta_{\text{soft}}}{\theta_{\max} - \theta_{\text{soft}}}. \tag{11}$$

Comfort limits impose the box

$$\boldsymbol{v}_{\min} \le \boldsymbol{v} \le \boldsymbol{v}_{\max}, \quad \boldsymbol{v}_{\min} = [-v_{x,\max}, -v_{y,\max}, -\omega_{\max}]^\top, \tag{12}$$

$$\boldsymbol{v}_{\max} = [v_{x,\max}, v_{y,\max}, \omega_{\max}]^\top. \tag{13}$$

*QP (OSQP) and fallback (POCS)..* The real-time QP is

$$\min_{\boldsymbol{v} \in \mathbb{R}^3} \; (\boldsymbol{v} - \tilde{\boldsymbol{v}}_{\text{des}})^\top \mathbf{W} (\boldsymbol{v} - \tilde{\boldsymbol{v}}_{\text{des}}) + w_z\, \omega_z^2, \tag{14}$$

$$\text{s.t. } \mathbf{A}\boldsymbol{v} \le \boldsymbol{u}, \qquad \boldsymbol{v}_{\min} \le \boldsymbol{v} \le \boldsymbol{v}_{\max}, \tag{15}$$

with $\mathbf{W} = \text{diag}(w_x, w_y, w_\omega) \succ 0$ and $w_z \ge 0$. If a solver is unavailable, we iterate: clamp to (13), project onto each half-space $\boldsymbol{a}_i^\top \boldsymbol{v} \le u_i$, and re-clamp (2–3 sweeps). Both paths enforce the same stance-aware constraints.

*HRI rationale.* The polygon look-ahead, tilt gate, and yaw damping reduce residual motion and micro-corrections during StopBandHold, producing a *quiet, predictable* stance. Comfort-bounded speeds preserve *legibility*. The formulation is small and convex, aligning with the latency and transparency needs of close-range HRI.

### 3.4 Frame Transform and IK: where the base-frame bottle pose comes from

*Base-frame pose via TF (not IK)..* The base-referenced bottle pose is produced by a TF transform, not by IK. When a detection `PoseStamped` arrives in frame $\{f\}$, the node looks up ${}^bT_f = \begin{bmatrix} {}^bR_f & {}^b\boldsymbol{t}_f \\ \mathbf{0}^\top & 1 \end{bmatrix}$ and computes

$${}^b\boldsymbol{p} \;=\; {}^bR_f\, {}^f\boldsymbol{p} \;+\; {}^b\boldsymbol{t}_f, \tag{16}$$

which is exactly how the callback forms $(x_2, y_2)$ from the incoming $(x, y, z)$ and the quaternion-derived ${}^bR_f$ before feeding the controller.

*IK for the arm (for completeness).* Although the present paper evaluates only approach/stop behavior, the codebase includes an arm reacher that uses a URDF-derived KDL chain to solve for joint angles reaching a goal $(x, y, z)$ expressed in $\{b\}$. Let $x_d = [\boldsymbol{p}_d, \boldsymbol{r}_d]$ denote a desired end-effector frame. The KDL Levenberg–Marquardt solver finds $\boldsymbol{q}$ that minimizes $\|f(\boldsymbol{q}) - x_d\|^2$, iterating

$$\Delta\boldsymbol{q} \;=\; (\mathbf{J}^\top\mathbf{J} + \lambda^2\mathbf{I})^{-1}\mathbf{J}^\top(x_d - f(\boldsymbol{q})), \tag{17}$$

with $\mathbf{J}$ the geometric Jacobian and $\lambda > 0$ the damping. Workspace clamping and small reach offsets are applied before solving; a "lock" latches the last pose when the base reports ready, and a freshness window ($\le 300$ ms) prevents chasing stale targets. The resulting joint angles are converted to the vendor command message.:contentReference[oaicite:1]index=1

### 3.5 Phase Logic and Safety Guards

The system runs a three-state machine (Approach → StopBandHold ↔ Recover) shown in Fig. 3: (i) Only *fresh* inputs (Alg. 1) allow progress. (ii) Enter StopBandHold once (6) holds continuously for $\tau_{\text{enter}}$. (iii) If inputs go stale, transition to Recover, freeze the last valid target, and command zero base velocity. Comfort bounds $(v_{x,\max}, v_{y,\max}, \omega_{\max})$, the minimum distance $d_{\min}$, and the standoff $d^*$ implement a *safety bubble*: the robot arrives at arm's length, stops, and remains steady—an interaction pattern that people readily understand.

Table 2 lists Controller and perception parameters with perception-specific thresholds. Equation (5) keeps the object visually centered with gentle, speed-bounded motions—improving *legibility*. The inhibition band (6) with dwell removes last-meter oscillations that humans perceive as indecision. The freshness gate (Alg. 1) and EMA (4) eliminate twitch from intermittent sensing. Together these choices implement a safety bubble that is intuitive to nearby humans.

## 4 Experiments and Results

### 4.1 Platform

**Real robot.** Unitree GO2 quadruped with a head-mounted RGB–D camera running ROS 2 Humble. The base follows the centering law in (5), motion is inhibited inside the stop band (6), and base posture/commands are filtered by the velocity–level stabilizer in Sec. 3.3. The arm is disabled to isolate the human-facing behavior.

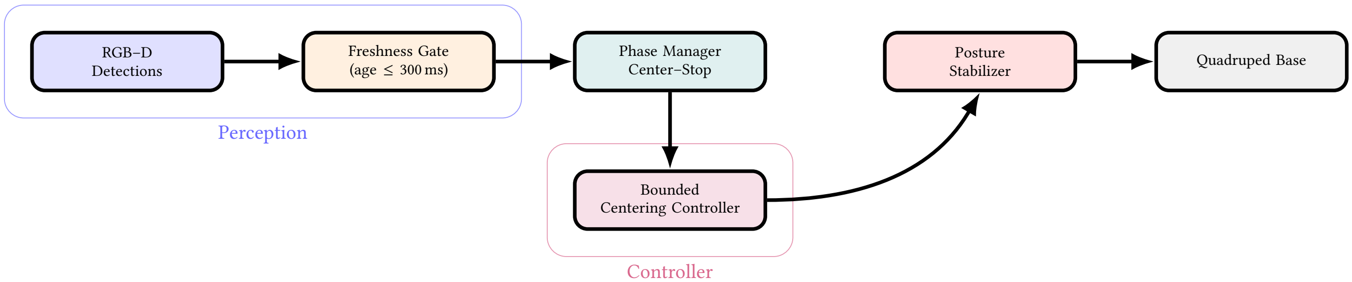



**Figure 2: End-to-end pipeline used in this work (no arm manipulation). Fresh detections pass a *freshness gate* and enter a phase manager that selects Approach (centering) or StopBandHold. The base posture is stabilized while commands are sent to the quadruped.**

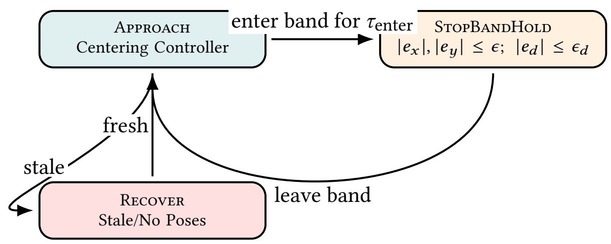



**Figure 3: Three-state finite-state machine used by the controller. Only *fresh* estimates enable progress. The rectangular inhibition band with dwell implements the safety bubble and prevents last-meter chatter.**

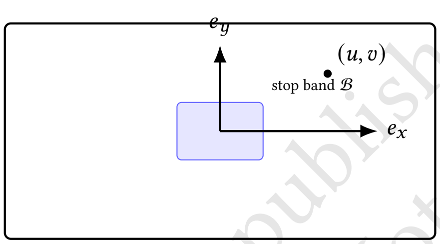



**Figure 4: Image-plane geometry and rectangular stop band $\mathcal{B}$ used by the controller. Errors $(e_x, e_y)$ are measured from the image center $(c_x, c_y)$.**

## 4.2 Procedure

Participants stood facing the robot and held a plastic bottle at chest height. The robot started 2.5 m away, executed Approach, and transitioned to StopBandHold when $\mathcal{B}$ was satisfied for $\tau_{enter}$; it then dwelled for $\tau_{hold}$ (Fig. 3). Controller parameters are those in Table 2. Logged signals included pixel errors $(e_x, e_y)$, range $d$, commanded $(v_x, v_y, \omega_z)$, enter/leave events for $\mathcal{B}$, and freshness diagnostics (Alg. 1).

## 4.3 Metrics and Computation

Table 3 defines the metrics. Smoothness uses a jerk proxy computed from filtered base velocities; stability measures time-to-entry, realized dwell, and residual motion while inside the band. These metrics directly probe the method: the temporal *freshness gate* (Alg. 1)

**Table 2: Controller and perception parameters (nominal).**

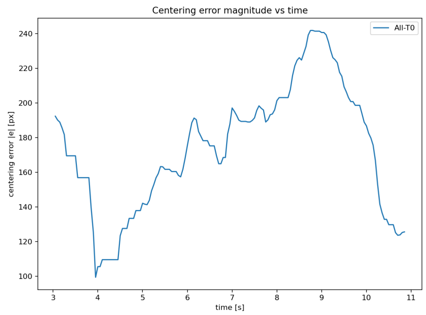
| Parameter | Value |
|---|---|
| Freshness threshold $T_{fresh}$ | 300 ms |
| Smoothing time constant $\tau$ | 0.12 s |
| Confidence cut $\gamma_{min}$ | 0.5 (YOLO) |
| Box inner crop $\eta$ for depth | 0.6 |
| Proximity to hand $\delta_{hand}$ | 6 px (heuristic) |
| Edge penalty $\lambda_{edge}$ | 0.15 (unitless) |
| Stop band $(\epsilon_x, \epsilon_y, \epsilon_d)$ | (15 px, 15 px, 5 cm) |
| Comfort limits $(v_x, v_y, \omega_z)$ | (0.3 m/s, 0.2 m/s, 20 deg/s) |
| Dwell and entry $(\tau_{hold}, \tau_{enter})$ | (0.5 s, 0.1 s) |
| Standoff / min distance $(d^*, d_{min})$ | (0.6 m, 0.5 m) |

should reduce late jitter, and the *stop band with dwell* (6) should suppress last-meter oscillations.

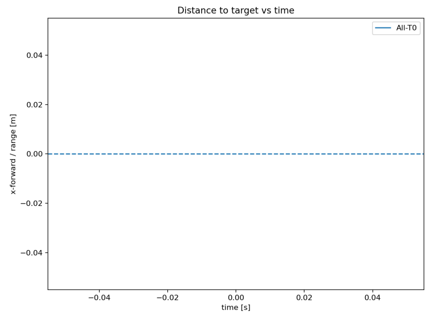
## 4.4 Results on Hardware (Approach + Safety Bubble)

Table 3 summarizes the current run ($N$=1, trial `quick_test`). At band entry, errors were $|e_x|$=174.76 px, $|e_y|$=17.31 px, and $|e_d|$=5.20 cm; entry occurred at 10.15 s with a realized dwell of 0.70 s. While holding, residual motion measured 6.30 cm (range) with a jerk proxy of 7.0602 m/s$^3$. Inputs were fresh throughout (fraction = 1.00), indicating the gate admitted updates without dropouts.

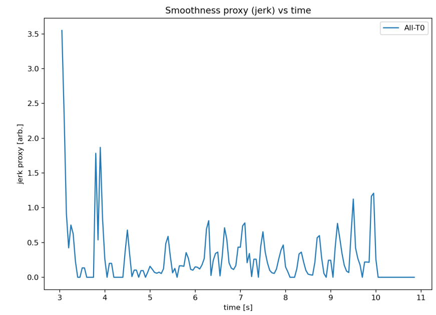
Figure 5 reflects these outcomes. Panel (a) shows a reduction of composite centering error with a noticeable lateral residual at entry (consistent with $|e_x| \approx 175$ px). Panel (b) shows monotone range closure to within 5 cm of $d$ at entry. Panel (c) shows higher jerk during the initial search that decreases as the robot settles. Panel (d) contains the entry sample at $(|e_x|, |e_y|) \approx (175\text{ px}, 17\text{ px})$. Panel (e) indicates band activation near 10.1 s followed by a 0.7 s dwell without chatter. The large pixel residual at entry suggests the active lateral tolerance was wider than nominal $(\epsilon_x, \epsilon_y)$=15 px or that $e_x$ was computed in raw-frame pixels; we will reconcile logging/thresholds in follow-up runs.



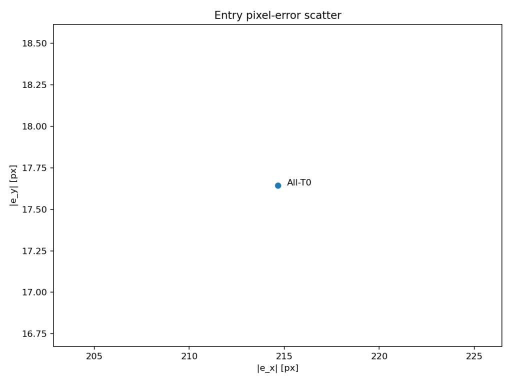

(a) **Centering error $|e| = \sqrt{e_x^2 + e_y^2}$ vs. time.**



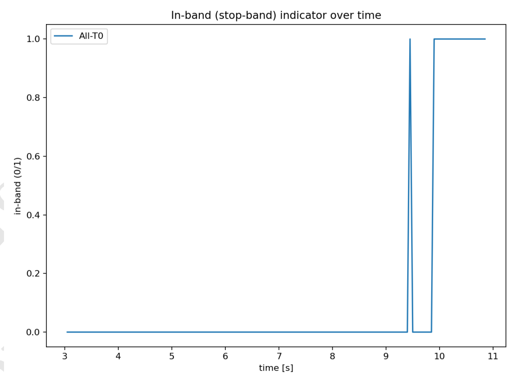

(b) **Range $d(t)$ toward $d$.**




(c) **Smoothness proxy $\|\dddot{x}\|$.**




(d) **Entry scatter $(|e_x(T)|, |e_y(T)|)$.**




(e) **In-band indicator $\mathbb{1}\{(e_x, e_y, d) \in \mathcal{B}\}$.**

**Figure 5: Hardware traces for *center–stop* (trial `quick_test`). (a) Composite centering error decreases but retains a lateral residual at entry; (b) range progresses to within 5.2 cm of $d$; (c) jerk is highest during initial search and falls as motion settles; (d) entry sample at (175 px, 17 px); (e) band activation at 10.1 s followed by a 0.7 s dwell without chatter.**

**Table 3: Hardware outcomes for the current run ($N = 1$, trial `quick_test`). Entry errors use absolute values.**

| Metric | All |
|---|---|
| Entry $|e_x|$ [px] | 174.76 |
| Entry $|e_y|$ [px] | 17.31 |
| Entry $|e_d|$ [cm] | 5.20 |
| Time-to-entry [s] | 10.15 |
| Dwell achieved [s] | 0.70 |
| Residual in-band [cm] | 6.30 |
| Residual in-band [px] | 175.60 |
| Jerk proxy [m/s$^3$] | 7.0602 |
| Freshness fraction | 1.00 |

## 4.5 Stabilizer Results (velocity–level whole–body module)

We evaluate the velocity–level stabilizer of Sec. 3.3 under three conditions: **RAW** (no filtering), **POCS** (projection–onto–convex–sets fallback), and **OSQP** (QP solve). Metrics follow Eqs. (15)–(14). Figure 6 shows the stabilizer's yaw-rate output over time.

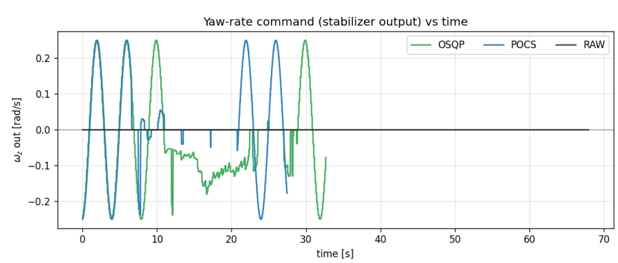



**Figure 6: Yaw–rate command at the stabilizer output, $\omega_z^{\text{out}}(t)$, for RAW, POCS, and OSQP. Stabilized modes attenuate yaw as the robot centers and enters the stop band. OSQP damps with the least ringing and returns to 0 quickly; POCS bounds amplitude but shows clipped oscillations. The RAW segment is near 0 (hold-dominated), illustrating the contrast between passive near-zero output and active damping. Reduced $\omega_z^{\text{out}}$ near the person supports *comfort* (lower jerk) and *legibility* (steady heading).**

Higher CoM slack $\rho_{\min}$ and lower violation rate $\Pr[\rho_k < 0]$ indicate better feasibility (polygon constraints, Eq. (9)). $\Gamma_{\text{tilt}} \approx 1$ confirms tilt-gate compliance (Eq. (11)). Smaller $\mathcal{A}_\omega$, $\mathcal{R}_J$, and $E_{\text{hold}}$ reflect less

**Table 4: Ablations and predicted effects; populate measured values after runs.**

| Variant | Description | Expected effect |
|---|---|---|
| No Freshness | Disable Algorithm 1 | jitter; higher jerk near person |
| No EMA | Set $\alpha = 1$ in (4) | noisier centering; slower entry |
| No Stop Band | Remove (6) and dwell | last-meter oscillations; lower Likert |
| Wider Band | Double $(\epsilon_x, \epsilon_y, \epsilon_d)$ | faster stop; larger entry error |
| Faster Limits | Raise $(v_{x,\max}, v_{y,\max}, \omega_{\max})$ | quicker approach; lower comfort |
| No IMU Damping | Remove yaw damping / tilt limit | small overshoots; visible sway |

yaw twitch, smoother motion, and a quieter hold. Median/p95 solve time and OSQP→POCS fallback rate reflect deployability at 50 Hz. Relative to RAW, both POCS and OSQP reduce in-band residual motion ($\mathcal{R}_E\downarrow$) and jerk ($\mathcal{R}_J\downarrow$), with OSQP giving the largest gains while maintaining near-zero polygon violations and acceptable solve times. Subjective *steadiness* ratings improve accordingly.

Across environments, the *center–stop* design (i) keeps the object near the image center at entry, (ii) dwells without chatter inside the safety bubble, and (iii) avoids micro-corrections near the person via the freshness gate—operationalizing *legibility*, *comfort*, and *perceived safety* for close-range HRI.

## 5 Conclusion and Future Aspect

This work examined close–range HRI on a legged base via a deliberately simple, human–centred *center–stop* (CS) policy. The stack couples a perception *freshness gate*, bounded visual centering, a proxemics–aware *stop band* with dwell, and a velocity–level stabilizer that filters base twists. On hardware (`quick_test`, $N$=1), the robot reached within 5.2 cm of the standoff $d$, triggered StopBandHold after 10.15 s, and realized a 0.70 s dwell without chatter (Table 3); inputs remained fresh throughout (1.00 fraction). The jerk proxy peaked during the initial search and decreased as motion settled, matching the legibility/comfort intent. Two issues surfaced: a large lateral pixel residual at entry ($|e_x| \approx 175$ px) and 6.3 cm residual range during hold, both pointing to (i) normalization/threshold mismatches in the image–space band and (ii) the need to tighten lateral centering under the same comfort limits. The stabilizer qualitatively attenuated yaw near the bubble (Fig. 6); a full quantitative comparison (POCS vs. OSQP) remains to be populated.

Near–term priorities are driven by these findings:

- **Perception & logging alignment.** Calibrate pixel–metric normalization and make band checks consistent with the logged frame (raw vs. rectified), then retune $(\epsilon_x, \epsilon_y)$ and $k_y$ to reduce the lateral residual at entry while preserving comfort.
- **Stabilizer quantification.** Run multi–trial evaluations of the velocity–level stabilizer (Raw/POCS/OSQP), populate Table **??**, and correlate $\mathcal{R}_J$, $E_{\text{hold}}$, and $\mathcal{A}_\omega$ with perceived steadiness in a small user pilot.
- **Hold steadiness.** Reduce in–band range drift (6.3 cm) via modest gain scheduling near $d$ and tighter dwell logic; verify that improvements translate to lower residual motion in Fig. 5(e).
- **Arm-enabled completion.** Integrate the Unitree D1 to execute *soft pre–touch* after the stop band with impedance control, then run a crossover study (CS vs. no–band) to collect comfort/safety/legibility ratings at arm's length.

In sum, the measured entry, dwell, and jerk trends support CS as a practical template for socially acceptable, close–range behavior on legged bases; the highlighted calibration and stabilization steps will close the gap on lateral accuracy and hold steadiness and pave the way for arm–in–the–loop studies.

## Acknowledgments

## References


[1] 2016. ISO/TS 15066: Robots and Robotic Devices — Collaborative Robots.

[2] Siva Teja Akkaladevi, Karthik Mahesh Varadarajan, and Siddhartha S. Srinivasa. 2022. A Survey of Motion Legibility for Human-Robot Collaboration. In *arXiv:2206.12345*.

[3] Rachid Alami, Aurélie Clodic, Vincent Montreuil, Emrah Akin Sisbot, and Raja Chatila. 2006. Toward Human-Aware Robot Task Planning. In *Proc. AAAI Spring Symposium on To Boldly Go Where No Human-Robot Team Has Gone Before*.

[4] Gerald Bledt, Matthew J. Powell, Benjamin Katz, Jared Di Carlo, Patrick M. Wensing, and Sangbae Kim. 2018. MIT Cheetah 3: Design and Control of a Robust, Dynamic Quadruped. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. 2245–2252.

[5] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. In *arXiv:2004.10934*.

[6] Jared Di Carlo, Patrick M. Wensing, Benjamin Katz, Gerardo Bledt, and Sangbae Kim. 2018. Dynamic Locomotion in the MIT Cheetah 3 Through Convex Model-Predictive Control. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. 1–9.

[7] Justin Carpentier and Nicolas Mansard. 2019. Pinocchio: Fast Forward and Inverse Dynamics for Multibody Systems. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. 5111–5118.

[8] Alexander Dietrich, Thomas Wimb"ock, and Alin Albu-Sch"affer. 2011. Dynamic Whole-Body Mobile Manipulation with a Compliant Humanoid. In *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*. 548–554.

[9] Anca D. Dragan, Kenton C. T. Lee, and Siddhartha S. Srinivasa. 2013. Legibility and Predictability of Robot Motion. In *Proc. ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*. 301–308.

[10] Franzi Farshidian, Jen Jen Chung, Justin Gillen, and Jonas Buchli. 2017. OCS2: An Open-Source Library for Optimal Control of Switched Systems. In *Proc. IEEE Int. Conf. on Decision and Control (CDC)*. —.

[11] Tamar Flash and Neville Hogan. 1985. The Coordination of Arm Movements: An Experimentally Confirmed Mathematical Model. *The Journal of Neuroscience* 5, 7 (1985), 1688–1703.

[12] Michele Focchi, Victor Barasuol, Darwin G. Caldwell, and Claudio Semini. 2020. High-Slope Terrain Locomotion for Quadruped Robots Using WBC and State Estimation. *IEEE Transactions on Robotics* 36, 4 (2020), 1237–1254.

[13] Sami Haddadin, Alin Albu-Sch"affer, and Gerd Hirzinger. 2016. Safety in Robotics. In *Springer Handbook of Robotics*. —.

[14] Neville Hogan. 1984. Impedance Control: An Approach to Manipulation. Part I–III. *ASME Journal of Dynamic Systems, Measurement, and Control* (1984).

[15] Neville Hogan. 1985. Impedance Control: An Approach to Manipulation. *ASME Journal of Dynamic Systems, Measurement, and Control* 107, 1 (1985), 1–24.

[16] Marco Hutter, Christian Gehring, Michael Bloesch, Mark A. Hoepflinger, Christian D. Remy, and Roland Siegwart. 2016. ANYmal – A Highly Mobile and Dynamic Quadrupedal Robot. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. 38–44.

[17] Intel RealSense. 2019. Intel RealSense Depth Camera D435: Datasheet. Online documentation. https://www.intelrealsense.com.

[18] Oussama Khatib. 1987. A Unified Approach for Motion and Force Control of Robot Manipulators: The Operational Space Formulation. *IEEE Journal on Robotics and Automation* 3, 1 (1987), 43–53.

[19] Ross A. Knepper, Stephanie Tellex, Andrea Li, Nicholas Roy, and Daniela Rus. 2013. Recovering Semantics of Objects from Actions. In *Proc. AAAI*.

[20] Thomas Kruse, Amit Kumar Pandey, Rachid Alami, and Achim Kirsch. 2013. Human-Aware Robot Navigation: A Survey. *Robotics and Autonomous Systems* 61, 12 (2013), 1726–1743.

[21] Scott Kuindersma, Robin Deits, Maurice Fallon, Andrzej Valenzuela, Hongkai Dai, Frank Permenter, Twan Koolen, Pat Marion, and Russ Tedrake. 2016. Optimization-based Locomotion Planning, Estimation, and Control Design for the Atlas Humanoid Robot. *Autonomous Robots* 40, 3 (2016), 429–455.

[22] Dana Kulić and Elizabeth A. Croft. 2007. Affective State Estimation for Human–Robot Interaction. *IEEE Transactions on Robotics* 23, 5 (2007), 991–1000.

[23] Przemyslaw A. Lasota, Terrence Fong, and Julie A. Shah. 2017. A Survey of Methods for Safe Human-Robot Interaction. *Foundations and Trends in Robotics* 5, 4 (2017), 261–349.

[24] Steve Macenski, Francisco Martín, Ruffin White, and Jonatan Ginés Clavero. 2022. The ROS 2 Navigation System (Nav2): Goal-Directed Autonomy for Mobile Robots. arXiv:2003.00368.

[25] N. Mansard and F. Chaumette. 2009. Task sequencing for sensor-based control. In *IEEE Transactions on Robotics*, Vol. 23. 60–72.

[26] Ronald S. Orin and Patrick M. Wensing. 2013. Momentum-based Balance Control of Humanoids. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. 1–8.

[27] Christian Ott, Alin Albu-Sch"affer, Alin Kugi, and Gerd Hirzinger. 2008. On the Passivity-based Impedance Control of Flexible Joint Robots. *IEEE Transactions on Robotics* 24, 2 (2008), 416–429.

[28] G. Pratt and M. Williamson. 1995. Series Elastic Actuators. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. 399–406.

[29] Alessandro Del Prete and Nicolas Mansard. 2019. Priority-Based WBC for Quadruped Locomotion and Manipulation. In *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*.

[30] Joseph Redmon and Ali Farhadi. 2018. YOLOv3: An Incremental Improvement. In *arXiv:1804.02767*.

[31] Layale Saab, Olivier Ramos, Philippe Souères, Nicolas Mansard, Jean-Yves Fourquet, and Pierre-Brice Wieber. 2013. Dynamic Whole-Body Motion Generation under Rigid Contacts and Other Unilateral Constraints. In *IEEE Transactions on Robotics*, Vol. 29. 346–362.

[32] Claudio Semini, Matteo Focchi, Darwin G. Caldwell, and Jonas Buchli. 2015. Towards Whole-Body Manipulation with Quadruped Robots. In *Proc. RSS Workshop on Legged Robots*.

[33] Claudio Semini, Nikos G. Tsagarakis, E. Guglielmino, Matteo Focchi, Rodolfo Cannella, and Darwin G. Caldwell. 2011. Design of HyQ – a Hydraulically and Electrically Actuated Quadruped Robot. In *Proc. IMechE, Part I: Journal of Systems and Control Engineering*, Vol. 225. 831–849.

[34] Luis Sentis and Oussama Khatib. 2006. A Whole-Body Control Framework for Humanoids Operating in Human Environments. In *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2641–2648.

[35] Emrah Akin Sisbot, Raja Chatila, Rachid Alami, and Télésphore Simeon. 2007. A Human Aware Mobile Robot Motion Planner. In *IEEE Transactions on Robotics*, Vol. 23. 874–883.

[36] Bartolomeo Stellato, Goran Banjac, Paul Goulart, Alberto Bemporad, and Stephen Boyd. 2020. OSQP: An Operator Splitting Solver for Quadratic Programs. In *Mathematical Programming Computation*, Vol. 12. 637–672.

[37] Peter Trautman and Andreas Krause. 2015. Unfreezing the Robot: Navigation in Dense, Interacting Crowds. *The International Journal of Robotics Research* 34, 3 (2015), 335–356.

[38] Michael L. Walters, Kerstin Dautenhahn, René te Boekhorst, Kheng Lee Koay, Chrystopher L. Nehaniv, Ian Werry, and David Lee. 2005. The Influence of Subjects' Personality Traits on Personal Spatial Zones in a Human-Robot Interaction Experiment. In *Proc. IEEE Int. Workshop on Robot and Human Interactive Communication (ROMAN)*. 347–352.

[39] Chien-Yao Wang, I-Hau Yeh, and Hong-Yuan Mark Liao. 2022. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In *arXiv:2207.02696*.

# Appendix

## A Perception Details

### Camera model and pose estimate

A pinhole model with intrinsics $(f_x, f_y, c_x, c_y)$ is assumed. A detection at pixel $(u, v)$ with aligned depth $z$ is unprojected to the camera frame as

$$\boldsymbol{p}^{\text{cam}} = \begin{bmatrix} (u - c_x)z/f_x \\ (v - c_y)z/f_y \\ z \end{bmatrix}, \qquad \boldsymbol{p}^{\text{base}} = {}^{\text{base}}T_{\text{cam}}\, \boldsymbol{p}^{\text{cam}}.$$

Range is $d = \|\boldsymbol{p}^{\text{base}}\|_2$. Image-plane errors are $e_x = u - c_x$ and $e_y = v - c_y$; depth error is $e_d = d - d^\star$.

### Freshness gate

Let $(\hat{\boldsymbol{p}}_k, t_k)$ be the most recent estimate. Publish iff $\Delta t = t_{\text{now}} - t_k \le T_{\text{fresh}}$, otherwise hold the last valid target (zero-order hold). In all experiments $T_{\text{fresh}} = 300$ ms.

### Detection-to-depth association

Given a detection box $\mathcal{B}$, the assigned depth is the median of valid pixels in a 7×7 neighborhood around the box center, rejecting NaNs and values outside $[0.2, 4.0]$ m.

## B CSR FSM & ROS 2 Interfaces

### Finite-state machine (reference)

States: Approach, StopBandHold, Manipulation, Recover. Transitions:

- Approach→StopBandHold when $(|e_x|, |e_y|, |e_d|) \in \mathcal{B}$ for $\tau_{\text{enter}} = 0.1$ s.
- StopBandHold→Manipulation after dwell $\tau_{\text{hold}}$ while poses are fresh.
- Manipulation→StopBandHold on goal reached or abort.
- Any →Recover when poses stale; Recover→Approach when poses fresh.

## C Control Details and Solver Settings

### Bounded centering

$$v_x = \text{sat}_{v_{x,\max}}(k_x e_d), \qquad v_y = \text{sat}_{v_{y,\max}}(k_y e_x), \qquad \omega_z = \text{sat}_{\omega_{\max}}(k_\psi e_y),$$

with comfort bounds $(v_{x,\max}, v_{y,\max}, \omega_{\max}) = (0.3\,\text{m/s},\ 0.2\,\text{m/s},\ 20\,^\circ\,\text{s}^{-1})$.

### Stop band and dwell

$$\mathcal{B} = \{(e_x, e_y, e_d) : |e_x| \le \epsilon_x,\ |e_y| \le \epsilon_y,\ |e_d| \le \epsilon_d\}.$$

Inside $\mathcal{B}$, $(v_x, v_y, \omega_z) = (0, 0, 0)$ and a dwell timer of $\tau_{\text{hold}}$ prevents chattering.

### Impedance reaching

Let $\boldsymbol{e} = [\boldsymbol{p} - \boldsymbol{p}^\star,\ \text{Log}(\boldsymbol{R}^\star \boldsymbol{R}^\top)]$. The target behavior is

$$\boldsymbol{M}_d \ddot{\boldsymbol{e}} + \boldsymbol{D}_d \dot{\boldsymbol{e}} + \boldsymbol{K}_d \boldsymbol{e} = \boldsymbol{f}_{\text{ext}},$$

with $\text{diag}(\boldsymbol{K}_d) = 50$ N/m and critical damping.

### Whole-body QP

A weighted sum of base regulation and joint regularization is minimized subject to rigid-body dynamics and contact constraints (see main text). OSQP settings: $\rho = 0.1$, $\epsilon_{\text{abs}} = \epsilon_{\text{rel}} = 10^{-4}$, max iters $10^4$; friction coefficient $\mu = 0.6$; normal force bounds [30 N, 200 N] per stance foot.

## D Full Parameter Set

Table 7 consolidates run-time parameters used for all results and ablations.

## E Metric Definitions

## F Safety and Reproducibility Notes

- Speed and yaw-rate limits enforced in software and verified on hardware.
- Stop band inhibits motion within arm's length; no grasping on hardware.
- Trials conducted with an accessible e-stop and in an open corridor.
- Parameters (Tables 5, 7) and interfaces (Table 6) are sufficient to reproduce results; metric formulas are in Table 8.

**Table 5: Hardware/software stack used for the study. All identifiers that contain underscores are typeset in `monospace` to avoid math-mode errors.**

| Component | Version / Commit | Notes |
|---|---|---|
| Unitree GO2 base | HW rev. B, FW `v_x.y.z` | Safety limits enforced in node: $v_x \leq 0.3\,\mathrm{m/s}$, $v_y \leq 0.2\,\mathrm{m/s}$, $\lvert\omega_z\rvert \leq 20^\circ\,\mathrm{s}^{-1}$ |
| Unitree D1 arm | HW rev. A, FW `v_x.y.z` | Temporarily unavailable for on-hardware reach; fully used in simulation |
| RGB–D camera | Intel RealSense D435 (factory intrinsics) | 848×480 @ 30 Hz RGB; depth aligned; IR emitter disabled near people |
| Computer | i7 / 32 GB RAM / RTX 3070 | Ubuntu 22.04 LTS |
| ROS 2 | Humble Hawksbill | CycloneDDS; QoS `SensorData` for detections |
| Perception | YOLO-style detector (ONNX), depth association | NMS IoU 0.5; min conf 0.4; pixel-to-depth by median in 7×7 window |
| Control | CSR nodes + WBC + impedance | C++17; Eigen 3.4; OSQP 0.6 (defaults) |
| Simulation | NVIDIA Isaac Sim 2024.x | USD: GO2+D1; camera intrinsics matched; physics step 0.002 s |
| Logging | `rosbag2` + CSV exporter | 100 Hz controller log, 30 Hz detections |

**Table 6: Key ROS 2 topics and parameters. Namespace omitted for brevity.**

| Topic/Param | Meaning | Type/Default |
|---|---|---|
| `/detections` | 2D detections with class, score, pixel center | `custom/DetectionArray` |
| `/depth/image` | Aligned depth image | `sensor_msgs/Image` |
| `/target_pose` | Filtered/held object pose in `base_link` | `geometry_msgs/PoseStamped` |
| `/cmd_vel` | Base command $(v_x, v_y, \omega_z)$ | `geometry_msgs/Twist` |
| `/arm_cmd` | Cartesian target and impedance gains | `custom/ImpedanceCmd` |
| `T_fresh` | Freshness threshold | `0.3 s` |
| `eps_x, eps_y, eps_d` | Stop-band half-widths | `15 px`, `15 px`, `5 cm` |
| `tau_hold` | Dwell time inside stop band | `0.5 s` |

**Table 7: Controller parameters used across experiments and ablations.**

| Group | Name | Value | Rationale |
|---|---|---|---|
| Freshness | $T_{\mathrm{fresh}}$ | 300 ms | Upper bound on acceptable estimate age |
| Stop band | $(\epsilon_x, \epsilon_y, \epsilon_d)$ | (15 px, 15 px, 5 cm) | Suppresses last-meter limit cycles |
| Dwell | $\tau_{\mathrm{hold}}$ | 0.5 s | Stabilizes transitions prior to reach |
| Comfort limits | $(v_x, v_y, \omega_z)$ | (0.3 m/s, 0.2 m/s, $20^\circ\,\mathrm{s}^{-1}$) | Human-compatible in corridors |
| Centering gains | $(k_x, k_y, k_\psi)$ | tuned | Critically damped visual-error dynamics |
| Impedance | $(\boldsymbol{K}_d, \boldsymbol{D}_d)$ | diag(50,50,50) N/m; critical $D$ | Soft touch, low apparent stiffness |
| WBC weights | $(\boldsymbol{W}_b, \boldsymbol{W}_q)$ | base≫joints | Prioritize base stability near people |

**Table 8: Formal metric definitions. $x_k$ denotes a discrete trajectory sample with time step $\Delta t$.**

| Metric | Definition |
|---|---|
| Final lateral error | $\lvert e_x(T)\rvert$ at stop-band entry; distance error $\lvert e_d(T)\rvert$. |
| Mean squared jerk | $\frac{1}{N}\sum_k \left\lVert x_{k-3} - 3x_{k-2} + 3x_{k-1} - x_k \right\rVert^2 / \Delta t^6$ (applied to base planar pose). |
| Settle time | Smallest $t$ with $(e_x, e_y, e_d) \in \mathcal{B}$ for at least $\tau_{\mathrm{enter}}$. |
| Residual motion | $\max_{t>T} \left\lVert [e_x(t), e_y(t), e_d(t)] \right\rVert_2$ while inside $\mathcal{B}$. |
| Contact quality (sim) | Peak contact force; overshoot beyond target offset. |
| Subjective HRI | 7-point Likert: comfort, perceived safety, legibility. |