CHASER: Robust Human Following on a Legged Robot Using Vision and Bluetooth

Julian Raheema[1,2], Connor Gag[1], Tyler Flar[1], Helena Bender[1], Matthew Zheng[1], Kurt Talke[2], Ryan Kastner[1], and Henrik Christensen[1]

***Abstract*— Human-following robots have achieved commercial adoption in structured indoor environments such as healthcare facilities, airports, and educational campuses, where environmental conditions are predictable and operational risk is low. However, their deployment in high-risk and unstructured domains, including firefighting, search and rescue, and defense operations remains limited. Existing platforms either rely on visual tracking, which struggles when the user leaves the camera field of view or when under low-light conditions. Radio-based followers provide reasonable accuracy, but suffer from poor resolution when using a single locator, signal interference, and limited applicability in certain environments. To overcome these limitations, we present a system that fuses a depth-camera perception pipeline with a single-antenna Bluetooth Angle-of-Arrival (AoA) direction estimator mounted on a quadruped robot. Our novel visual tracker runs a YOLOv8n detector, extracts multimodal embeddings (RGB, depth, pose, colored arm-sleeve histograms) and re-identifies the operator at 5 Hz on a CPU-only platform. When visual contact is challenged, the AoA module supplies a heading cue that steers the robot back into view. Experiments were conducted in unknown, unstructured, indoor and outdoor environments. Using a Spot legged robot, we show that the hybrid approach reduces tracking interruptions by 70% compared with vision-only, while maintaining a lateral position error of 0.54m (RMSE) and a heading error of 0.31rad. The result is a hybrid system capable of robust real-time user following across diverse environments. We call our system *CHASER* (Collaborative Helper Autonomous System for Exploration Robots), a wearable-enabled multimodal follower designed for field deployment.**

I. INTRODUCTION

Human-following robots have the potential to substantially improve daily life, professional workflows, and emergency response operations by providing mobile assistance and reducing cognitive load on operators. Prior studies have demonstrated their value in domains such as healthcare [1], airport baggage handling [2], and disaster response [3], where autonomous following enables more efficient task execution and improved human-robot teaming. Autonomous robot followers have strong potential in high-risk operations, including bomb disposal, hazardous material inspection, and delivery of medical supplies during complex missions.

As of today, most commercial robot followers have focused on wheeled platforms for structured environments. For example, systems such as Skydio have demonstrated visual tracking autonomously in aerial drones for tasks such as inspection and cinematography [4]. In ground robotics, robot followers have been researched and deployed in environments like airports, where wheeled robots assist passengers [5]. The Unitree wireless vector positioning system enables legged robots to follow humans outdoors, but its accuracy degrades in cluttered environments due to multipath interference and occlusions [6], [7]. Although autonomous follower systems have been explored in structured environments, limited research has focused on user-following robots operating in unstructured or disaster-response scenarios [8]. Current search-and-rescue robots are often teleoperated, requiring operators to divide attention between robot control and mission-critical tasks such as situational awareness and target identification.

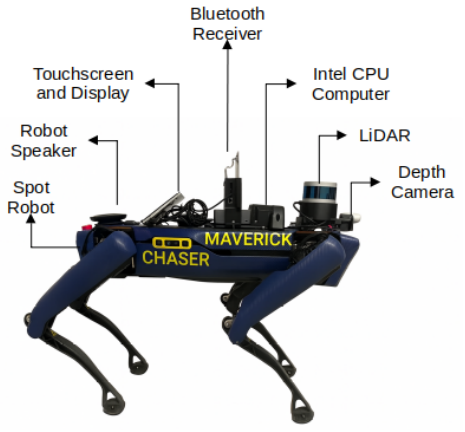

Fig. 1. CHASER: An Autonomous Human-Following Robot System

An autonomous robot follower with sensing and payload capabilities can improve mission efficiency by reducing operator workload. Legged robots are especially advantageous in rough or debris-filled terrain where wheeled and tracked systems struggle. Legged human-following robots can support responders by transporting sensing payloads, communication equipment, and mission-critical tools through challenging terrain while reducing operator workload during field operations. Such systems can enhance situational awareness, extend operational capabilities, and allow operators to remain focused on critical mission objectives.

Our system, shown in Fig. 1, combines depth camera and Bluetooth sensing to enable human-following on a legged robot platform, with control provided through a smartwatch interface for indoor and outdoor field use. Vision is used for metric tracking when the operator is in view, while Bluetooth Angle-of-Arrival (AoA) provides a directional cue to support recovery when visual tracking is degraded or lost.

[1]Computer Science and Engineering, University of California San Diego, 9500 Gilman Drive, La Jolla, CA

[2] Naval Information Warfare Center Pacific, 49275 Electron Dr. San Diego, CA, USA

Contributions: We present a deployable human-following legged robot system designed for robust operation under practical constraints, including CPU-only computation, lighting variation, and partial occlusions.

i. A **wearable smartwatch command-and-safety interface** integrated with ROS for rapid mode switching.
ii. A **CPU-only multimodal re-identification pipeline** (RGB, depth, pose, and forearm appearance) enabling reliable target reacquisition at ≈5 Hz.
iii. A **single-locator Bluetooth AoA heading module** providing a high-rate (≈37–50 Hz) directional cue for recovery.
iv. A **practical multimodal arbitration strategy** that prioritizes vision for metric following while leveraging Bluetooth AoA for reorientation, validated on a full-size quadruped in indoor and outdoor environments.

II. DESIGN REQUIREMENTS AND SYSTEM GOALS

The system is guided by the following requirements:

- **Human-centered control and safety** via a wearable interface with rapid mode switching and emergency stop.
- **CPU-only real-time operation** without reliance on a discrete GPU.
- **Graceful recovery** under lighting changes and partial occlusions.
- **Minimal infrastructure**, using a single robot-mounted Bluetooth locator and one user-carried tag.
- **Deployability**, producing standard ROS velocity commands for platform portability.

III. RELATED WORK

This work extends an earlier graduate project study on wearable-guided human-following robots at UC San Diego.

A. Operator control Unit: Smartwatch

Recent progress in wearable devices has motivated a variety of approaches for robot control and human-robot interaction (HRI). For instance, iRoCo [9] combines smartwatch and smartphone sensing with Kalman Filters to estimate arm motion for robot teleoperation. Although effective, these methods often depend on continuous motion estimation, calibration procedures, and computationally intensive processing, which may reduce robustness in constrained or noisy environments. WearMoCap [10] further expands multimodal body-motion tracking using wearable smart devices positioned at different body locations, but challenges such as sensor drift, and user comfort remain. In addition, the approach proposed in [11] estimates arm pose using smartwatch data and machine learning techniques alone, achieving promising accuracy while introducing increased algorithmic complexity and real-time processing demands.

B. Bluetooth-Based Human Following

Bluetooth-based positioning and localization has emerged as a lightweight and energy-efficient solution for tracking subsystems, especially in indoor and GPS-denied environments where traditional localization methods are limited. The proposed directional localization framework relies on a wearable Bluetooth transmitter carried by the user and a robot-mounted AoA receiver array that estimates relative heading information from received wireless signals. Received Signal Strength Indicator (RSSI) is used to estimate tag distance through the power level of a received wireless signal, typically expressed in dBm. It is used in Wi-Fi, Bluetooth, and other radio systems to estimate link quality or approximate distance between devices. Typically, at least 2 locators are required for accurate pose using Bluetooth, Wi-Fi, or similar radio-based positioning systems [12] [13]. Prior studies that rely on infrastructure to localize the robot using multiple Bluetooth locators, such as [14], require additional hardware and a fixed location to estimate the pose of the Bluetooth transmitter. Integrating Bluetooth with additional sensors can improve accuracy, as demonstrated in [15], where the authors combined RSSI with smartphone IMU data to extend the robot follower's range to approximately 4 meters, achieving a positional error with a standard deviation of ±70 cm. Sátán and Tóth [16] also investigated indoor proximity localization using filtered RSSI measurements combined with log-distance path loss models. Their approach achieved practical room-level localization accuracy, though it lacked precise directional or angular estimation. However, due to the 500 ms transmission interval of the beacons, reduced accuracy during movement (which could be improved with dead reckoning), and signal attenuation caused by the user's body orientation; additionally, the system may provide unrealistic results in very close ranges, requiring special handling. More recent approaches utilize Ultra-Wideband (UWB) tracking subsystems, such as smart-stroller [17], which achieve centimeter-level accuracy and strong resistance to interference, but they come with increased hardware complexity and have so far been limited to indoor testing. One potential source of inaccuracies occurs when the sensor planes are misaligned due to bouncing in an unstructured environment.

C. Vision-Based Human Following

The Vision-based robot follower system employs a depth or stereo camera to detect and track the target. One study using vision system cart-following system using CMUcam5 color tracking was developed for wheelchair assistance [18]. However, its reliance on color makes it highly sensitive to lighting, background interference, and limited field of view, reducing tracking reliability. An indoor deep-learning robot follower using Single Shot MultiBox Detector (SSD) with color features employed a state machine for recovery but was limited by similar clothing and lighting changes [19]. A more recent Vision-based robot follower [20] presents a Kinect-based human-following robot using SURF features, Kalman filtering, and visual servo control for tracking. It

TABLE I

COMPARISON OF DEEPSORT AND CUSTOM RE-ID PIPELINE

Feature	DeepSORT	Ours
Modality	RGB only	RGB + Depth + Pose + Color
ID Switches	Frequent in crowds	Reduced with fusion
Computation	GPU preferred	CPU, 5 Hz
Flexibility	Fixed features	Modular, extensible
Training	Offline Hours	Online 15 Seconds

performs well indoors but struggles with pose changes and occlusions. Person re-identification assigns a unique ID to an individual, enabling the system to recognize them upon exiting and reentering the camera's field of view. It has been studied extensively across multiple modalities, including RGB, depth, skeleton, and infrared [21], [22]. Fusion strategies are typically categorized into feature-level and score-level approaches [23]. One limitation is that these methods require extensive labeled data and training resources.

Alternative approaches, such as YOLOv8n with DeepSort [24] and MobileNetV2 [25], showed limitations in handling occlusions, frequent ID reassignment, and computational constraints on non-GPU platforms. These shortcomings ultimately motivated the development of our custom re-identification framework, which prioritizes adaptability, efficiency, and reliable performance across both indoor and outdoor environments (shown in Table I).

IV. SYSTEM ARCHITECTURE

A. Hardware Architecture and System Components

The system uses a set of hardware components to enable the robot to track and follow a leader. We use the Boston Dynamics' Spot quadruped robot as a robot template, known for its agility, stability, and mobility across varied terrain [26] (see Fig 2). A wearable smartwatch interface was used to support remote command execution, mode switching, and operator interaction during field testing. For Vision-based human following, an Intel RealSense D455 depth camera [27] was integrated to enable 3D user detection and tracking using computer vision techniques. The Bluetooth-based following was implemented using the EFR32BG22 Bluetooth Dual Polarized Antenna Array Pro Kit and Thunderboard Kit from Silicon Labs [28]. In this setup, the user carries the Thunderboard device while the EFR32BG22 antenna array is mounted on the robot, allowing estimation of the user's relative position and orientation using Bluetooth signal measurements. All onboard processing is performed on an Intel NUC Mini-PC equipped with a 22-core Intel Ultra 9 processor, which handles sensor fusion, autonomy, SLAM, and robot follower algorithms in real time while interfacing directly with the Spot robot control system. To improve vision tracking robustness, the user wore red and blue armbands as visual markers.

B. System Architecture

The overall system architecture, illustrated in Fig. 3, combines multiple hardware and software components to support

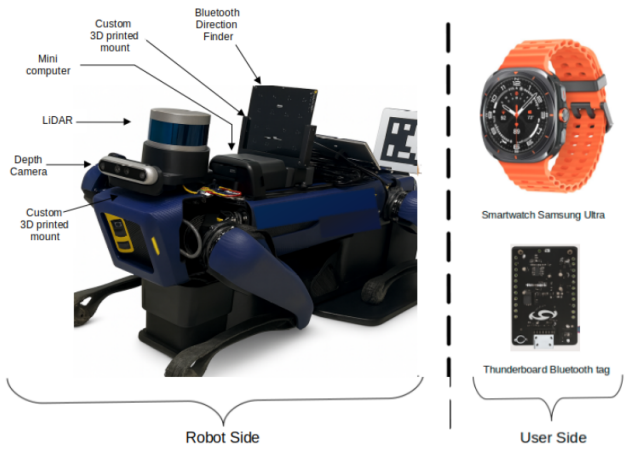

Fig. 2. Boston Dynamics' Spot equipped with sensors and an autonomy computer; the user carries a Galaxy Watch Ultra and Bluetooth beacon for control and directional tracking.

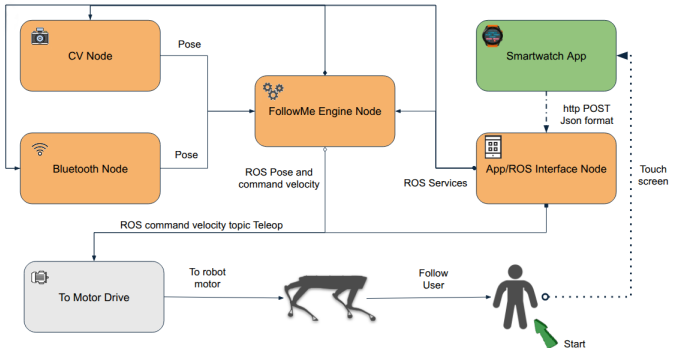

Fig. 3. CHASER system architecture integrating computer vision, Bluetooth AoA localization, and smartwatch control inputs through the follower engine node to generate autonomous robot-following commands.

coordinated interaction between sensing systems, user interfaces, and robotic control modules. System integration and inter-process communication were implemented within a Robot Operating System (ROS) [29] framework running on Ubuntu Linux, enabling modular interaction between sensing, control, and navigation components. A custom Android smartwatch application was developed using Android Studio to provide a portable interface for issuing motion commands and activating tracking modes remotely. The smartwatch communicates wirelessly with the robot, enabling intuitive human-robot interaction in the field. This wearable interface allows operators to interact with the robot while maintaining mobility and situational awareness. For Bluetooth-based localization, the EFR32BG22 Bluetooth Direction Finding Kit was programmed using Silicon Labs Simplicity Studio. This setup enabled implementation of Bluetooth Angle-of-Arrival (AoA) techniques for estimating the user's relative direction and position. The proposed framework extends prior work on autonomous exploration, SLAM [30], and navigation for quadruped robotic systems [31]. The architecture is modular and robot-agnostic, allowing deployment on ROS-compatible robotic platforms that support the

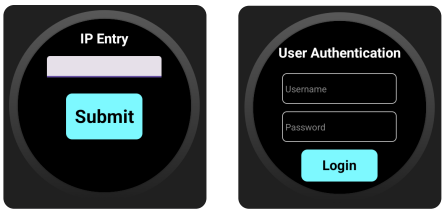

Fig. 4. Smartwatch interface for robot connection, showing IP address entry (left) and followed by user authentication step (right).

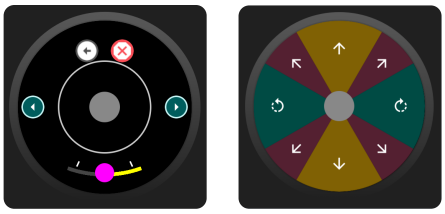

Fig. 5. Smartwatch teleoperation interface: initial state (left) and active state (right) with enlarged control area and hidden nonessential buttons.

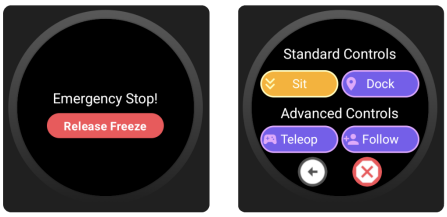

Fig. 6. Smartwatch interface showing the emergency stop function (left) and the command menu with standard and advanced robot controls (right).

required hardware interfaces and standard velocity command inputs.

V. METHODOLOGY

A. Implementation

The operator communicates with the robot using a smartwatch application interface that transmits HTTP requests. A server running on a specified IP address receives these commands and either invokes ROS services or publishes messages to designated ROS topics. The framework allows users to choose between Vision, Bluetooth, or combined tracking modes; at present, Bluetooth is used primarily for directional estimation, while distance estimation is planned for future pose tracking capabilities. After estimating the target location, a proportional controller directs the robot toward the user with the desired heading and speed.

B. Operator Control Unit

The smartwatch application was designed with an emphasis on intuitive interaction and accessibility. The interface incorporates user-friendly layouts and color-aware design considerations to improve usability across different operational conditions. Upon launching the application on the smartwatch, the user is first presented with an IP configuration page for robot selection, followed by an authentication step shown in Fig. 4. The interface provides basic robot commands such as *sit*, *stand*, *dock*, and *undock*, along with advanced operational modes including *teleoperation* and *follow* modes. In teleoperation mode, the user accesses a dedicated control screen that enables manual robot navigation and rotational control through a virtual joystick interface. The joystick interface introduces a single-finger control design that combines translational and rotational motion into a unified input scheme. Shown in Fig. 5, the interface supports forward/backward, diagonal, and rotational movement. In follow mode, the user is directed to a tracking selection page with multiple options, including Apriltag [32], Computer Vision (CV), and Bluetooth (BT) tracking. The operator can enable one or multiple tracking modules for autonomous following. An Emergency Stop (E-Stop) option, shown in Fig. 6, is available when unsafe robot behavior is detected. To enhance safety, we implemented the E-Stop strategy to ensure human judgment remains prioritized over autonomous behavior.

C. Bluetooth-Based Following Module

1) System Concept: The Bluetooth-based localization subsystem utilizes a Silicon Labs BG22 Bluetooth Dual Polarized Antenna Array Pro Kit (BRD4191A) configured with a 4×4 dual-polarized URA antenna array to detect EFR32BG22 Thunderboard tags and estimate the user's relative azimuth and elevation angles. The smartwatch's Bluetooth was intended as a tag, but since consumer devices hide the Constant Tone Extension (CTE) needed for direction finding, we emulated smartwatch tags using EFR32BG22 Thunderboard Kits (BRD4184A), which provide firmware access and comparable over-the-air behavior. Each tag periodically appended a CTE to advertising packets, enabling the locator to estimate directions at 50 Hz. These measurements are relayed to the robot for real-time tracking and following.

2) Angle-of-Arrival (AoA) Measurement: The in-phase and quadrature (IQ) samples capture both the amplitude and phase of the received signal at each antenna element. When collected simultaneously across a Uniform Rectangular Array (URA), these samples form a spatial snapshot of the signal wavefront, enabling techniques such as beamforming and direction-of-arrival estimation. Following carrier-frequency compensation, the processed signal snapshot is provided to a MUltiple SIgnal Classification (MUSIC) algorithm for angle estimation [33]. MUSIC is a high-resolution algorithm for estimating angle of arrival (AoA). After the locator captures IQ samples during the Bluetooth LE (CTE), it forms a covariance matrix from those samples. The algorithm then performs an eigen-decomposition to separate the signal and noise subspaces. Finally, it scans candidate angles to compute a MUSIC, producing θ (azimuth) and ϕ (elevation) [33]. The estimated (θ), (ϕ), and range ρ form a spherical coordinate representation in the locator frame, with angular uncertainties of $\pm 1.97°$ in azimuth and $\pm 4.22°$ in elevation. This is converted to Cartesian coordinates using

$$x = \rho \sin\phi \cos\theta, \tag{1}$$
$$y = \rho \sin\phi \sin\theta, \tag{2}$$
$$z = \rho \cos\phi. \tag{3}$$

Our human-following algorithm uses the Bluetooth lateral measurement $y \in [-1, 1]$ (ignoring x and z) to compute a raw yaw error, shown in Eq. (4). The signal is processed through a dead-zone and hysteresis near zero to reduce limit cycling. A first-order Exponential Moving Average (EMA) filter is then applied to smooth the error, as defined in Eq. (5). Derivative damping is introduced according to Eq. (6), while the pre-saturation heading command follows the PD control law in Eq. (7). The scaling function is defined in Eq. (8). Finally, the heading command is saturated and debiased using Eq. (9), with a small deadband applied near zero.

$$\alpha_{\text{raw}} = \operatorname{atan2}(y, 1) \tag{4}$$
$$\alpha_{\text{ema}}[k] = \beta\, \alpha_{\text{use}}[k] + (1 - \beta)\, \alpha_{\text{ema}}[k-1] \tag{5}$$
$$\dot{\alpha}[k] = \frac{\alpha_{\text{ema}}[k] - \alpha_{\text{ema}}[k-1]}{\Delta t} \tag{6}$$
$$\omega^*[k] = k_p^{\text{bt}}\, s(|y|)\, \alpha_{\text{ema}}[k] - k_d^{\text{bt}}\, \dot{\alpha}[k] \tag{7}$$
$$s(r) = \max(r_{\min}, \min(1, r)) \tag{8}$$
$$\omega[k] = \text{sat}_{[-\omega_{\max}, \omega_{\max}]}\left(\omega^*[k]\right) \tag{9}$$

3) Range Estimation from RSSI: Experimental evaluation showed that RSSI-derived distance estimates were highly unstable under real-world conditions, frequently exhibiting severe discontinuities and unreliable range behavior. As a result, Bluetooth ranging was excluded from the final follower controller, while depth sensing remained the primary source for metric distance estimation.

D. Vision-Based Following Module

The perception framework combines lightweight person detection using YOLOv8n [34], multimodal feature extraction using OSNet for RGB and depth [35], MediaPipe for skeleton tracking [36], and a forearm color histogram for appearance cues under CPU-only constraints. To enhance temporal consistency and robustness, a Kalman filter was employed to predict future positions, restrict search regions, and effectively mitigate occlusion effects [37], [38]. The proposed framework focuses on tracking a single selected user within the robot's observable scene, as illustrated in Fig. 7. A key component of this approach is the Acquisition phase shown in Fig. 8, which is responsible for initializing and registering the target individual. Prior to autonomous tracking, the operator performs a short enrollment sequence in front of the robot to initialize appearance and pose descriptors across multiple sensing modalities. Only frames containing a single detected individual are retained to prevent ambiguous identity associations during feature enrollment. For each verified frame, feature extractors generate multi-modal embeddings, forming a dataset S as ground truth for future comparisons. After enrollment, the system enters the Tracking stage shown in Fig. 9, where the target is continuously re-identified in incoming frames. For every incoming frame, candidate detections are evaluated against the enrolled operator profile using modality-specific descriptors derived from appearance, depth, skeletal pose, and forearm color information encoded into a 512-dimensional vector. MediaPipe Pose extracts skeletal body-ratio features and Forearm regions are also used to compute color histogram features.

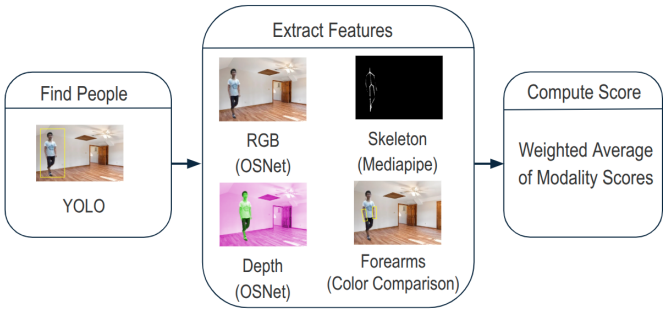

Fig. 7. Multimodal tracking pipeline using YOLO, RGB, depth, pose, and forearm color features for weighted identity matching.

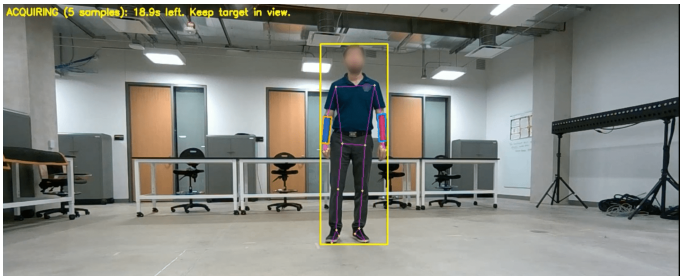

Fig. 8. Acquisition phase of the vision-based follower system, illustrating person detection for target enrollment and re-identification.

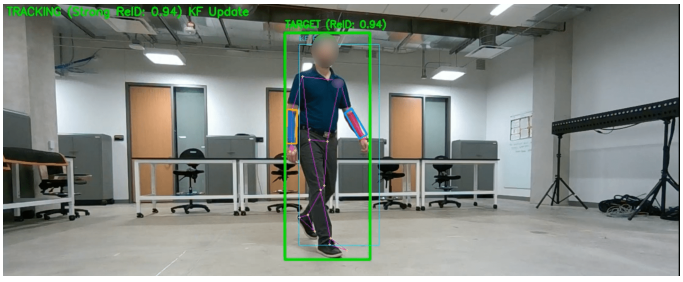

Fig. 9. Real-time target re-identification and tracking.

Once feature extraction is completed, Kalman filtering is applied to reduce candidate uncertainty before performing similarity evaluation using fused multi-modal weighting. Target association is performed through weighted multimodal score fusion, where similarity measurements from each sensing modality are aggregated to determine the most likely target identity. This matching process for a query person Q is

expressed by the following equation:

$$\text{Score}(Q,S) = \frac{1}{\text{length}(M)} \sum_{m \in M} w_m \cdot \left(\max_{s_m \in S_m} \text{cosine\_similarity}(f_{m,Q}, f_{m,s}) \right) \tag{10}$$

In this formulation, M represents the collection of active sensing modalities, while w_m denotes the weighting factor assigned to modality m. The set S_m contains the enrolled feature samples associated with modality m, where $f_{m,Q}$ corresponds to the query feature vector and $f_{m,s}$ represents an enrolled sample feature. The individual with the highest similarity score exceeding an adaptive re-identification threshold is selected as the target. This threshold is determined from score distributions computed during the acquisition phase, where each enrolled sample is evaluated against the remaining samples. After successful target association, the Kalman filter state is updated and the estimated 3D target coordinates are published to the navigation system. After the target is successfully re-identified, the system converts the detected 2D image location into a 3D position within the robot reference frame. This transformation uses the camera intrinsic parameters together with aligned depth measurements to estimate real-world coordinates. The target position is obtained from the center of the detected bounding box, while the depth value is computed using the median depth within the corresponding region. These measurements are then used to estimate the target's (X, Y, Z) position relative to the camera, allowing the robot to follow the user in 3D space.

E. Robot Follower Control and Navigation

As illustrated in Algorithm 1, enabling the Vision-based follower activates the motion engine and initiates communication with the computer vision (CV) module to acquire target features for tracking. The user's position is estimated based on the selected tracking mode, and once the target is detected, a lightweight proportional controller regulates the robot's heading and velocity relative to the user location. Motion commands are continuously updated using real-time sensor measurements, allowing the robot to maintain tracking, and avoiding obstacles using built-in Spot cameras.

In Bluetooth-only mode, the robot rotates to face the user's tag, as distance estimation is not yet implemented. In the combined Vision+Bluetooth mode, computer vision serves as the primary tracker, while Bluetooth acts as a fallback: if the user leaves the camera frame, the robot orients toward the tag and resumes Vision-based following once the user re-enters the frame.

$$v = k_d \cdot e_d \tag{11}$$

$$\omega = k_\theta \cdot e_\theta \tag{12}$$

The robot motion is controlled using a proportional controller, where the linear velocity is defined as $v = k_d \cdot e_d$ and the angular velocity as $\omega = k_\theta \cdot e_\theta$. In this formulation, e_d represents the distance error, e_θ denotes the heading error, and k_d and k_θ correspond to the proportional gains for distance and heading control. Obstacle avoidance and local collision prevention are handled using Spot's onboard perception and navigation stack during follower operation.

Algorithm 1: Robot Motion Engine Follower

Input: CV/BT poses, params $k_\rho, k_\alpha, k_\alpha^{bt}, k_d^{bt}$, limits $v_{\max}, \omega_{\max}$
Output: Command Velocity (v, ω)
foreach *tick at 30 Hz* **do**
  if *CV available* **then**
    $e_d \leftarrow \|p_{cv}\| - d^*$
    $v^\star \leftarrow \text{clip}(k_\rho e_d,\ -v_{\max}, v_{\max})$
  else
    $v^\star \leftarrow 0$
  if *CV available* **then**
    $\alpha \leftarrow \text{atan2}(y_{cv}, x_{cv})$
    $\omega^\star \leftarrow \text{clip}(k_\alpha \alpha,\ -\omega_{\max}, \omega_{\max})$
  else
    if *BT available* **then**
      $\alpha \leftarrow$ hysteresis+EMA(y_{bt})
      $\omega^\star \leftarrow \text{clip}(k_\alpha^{bt}\alpha - k_d^{bt}\dot{\alpha},\ -\omega_{\max}, \omega_{\max})$
    else
      $\omega^\star \leftarrow 0$
  $v \leftarrow \text{ramp}(v_{last}, v^\star)$; $\omega \leftarrow \text{ramp}(\omega_{last}, \omega^\star)$
  publish (v, ω);

Fig. 10. Left: outdoor route; Middle: robot following on gravel (vision only); Right: indoor following (vision+armbands+Bluetooth).

VI. TESTING SETUP

Experiments were conducted in both outdoor and indoor environments (Fig. 10). Outdoor tests covered gravel, dirt, forest, and paved surfaces under direct sunlight (≈100,000 lux) and shaded conditions (≈20,000 lux), with a total travel distance of ≈156 m. Indoor tests were performed in office and warehouse spaces under lighting levels between ≈210 and 1,220 lux, with a total distance of ≈154 m. A single target user was used throughout the experiments. Each sensing configuration and lighting condition was evaluated over three independent trials and averaged. The evaluated modes included Vision-only (V), Vision + Armbands (V+AB), and Vision + Armbands + Bluetooth (V+AB+BT). In two-person scenarios, an additional distractor user was introduced to evaluate re-identification robustness and false-follow behavior. Each trial began with a 15 s target enrollment phase, followed by autonomous tracking and following. Performance metrics included tracking interruptions (#stops), mis-follow events (#distracts), trajectory accuracy, heading error, recovery time after vision loss, and leader-follower separation distance. Teleoperation experiments were

TABLE II

ROBOT FOLLOWER LOST TRACKING (#STOPS) ACROSS ENVIRONMENTS AND SENSING MODES (1-PERSON SCENARIOS). V: VISION-ONLY, V+AB: VISION+ARMBANDS, V+AB+BT: VISION+ARMBANDS+BLUETOOTH.

Env/Light (in LUX)	**V** **#stops**	**V+AB** **#stops**	**V+AB+BT** **#stops**
Indoor (>220)	6.3	4.3	**3.3**
Indoor (<220)	6.6	8.6	**2.3**
Outdoor (>95K)	9.3	**3.3**	4.6
Outdoor (<10K)	1.7	0.7	**0.3**

also repeated three times and averaged for comparison against autonomous follower performance.

VII. RESULTS

The robot was configured as outlined in Section VI, and all experiments were carried out accordingly. The baseline time to complete the route at a slow walking pace, without the robot, was **2 min 10 s**. With the robot, completion times varied based on lighting conditions and the number of stops. For outdoor environments under **10,000 lux**, the fastest run (best) was **2 min 16 s** with zero interruptions using the combination of vision, armbands, and Bluetooth. The slowest run (worst) was **10 min 02 s** with ten stops under outdoor lighting above **95,000** lux using vision only. In the most extreme case, the robot failed to detect and follow the leader, and the route could not be completed. For indoor environments below **220 lux**, completion times ranged from **2 min 30 s** (best) when using the vision, armbands, and Bluetooth to **11 min 06 s** (worst) using vision only. Under lighting conditions above **220 lux**, the robot achieved a best time of **2 min 29 s** and a worst time of **4 min 30 s**. We also had some cases where the robot struggled to follow the leader when the lighting conditions changed drastically. Teleoperation results were averaged over three runs, and as shown in Table IV, the autonomous follower produced smoother but slower motion than teleoperation while maintaining moderate tracking accuracy (0.54 m RMSE position and 0.31 rad RMSE heading), as shown in Fig. 11. Additionally, we report the discrete Fréchet[39] distance as a leash metric capturing the worst-case separation while preserving temporal ordering. The follower maintained a Fréchet distance average of 1.91 m and a mean leader–follower separation of 1.15 m, indicating consistent proximity during traversal. Finally, in a range test using vision, armbands, and Bluetooth, the robot successfully followed the user for approximately 804 meters with only three stops under the Vision–Bluetooth configuration.

VIII. LIMITATIONS

The system prioritizes robustness and human-centered control, reducing computational complexity, avoiding sensor drift, and improving field usability. The Vision-based follower is sensitive to illumination, with degraded performance in low light and during abrupt lighting transitions. The Bluetooth follower provides a lightweight directional cue that supports recovery when vision is degraded or out of view. In our single-locator,

TABLE III

ROBOT FOLLOWER LOST TRACKING (#STOPS) AND MIS-FOLLOW (#DISTRACTS) FOR 2-PERSON SCENARIOS.

Env/Light (in LUX)	**V**		**V+AB**		**V+AB+BT**	
	#stops	**#distracts**	**#stops**	**#distracts**	**#stops**	**#distracts**
Indoor (>220)	7.6	2.0	6.3	0.0	**2.6**	0.0
Indoor (<220)	4.0	3.3	3.6	0.0	**1.3**	0.0
Outdoor (>95K)	6.6	5.6	**4.6**	4.0	6.5	3.5
Outdoor (<10K)	12.0	2.6	4.3	0.0	**1.6**	0.0

TABLE IV

TELEOP VS. FOLLOWER PERFORMANCE AND TRACKING ERRORS (MEAN ± STD OVER 3 RUNS).

Metric	**Teleop**	**Follower**
Duration [s]	71.6 ± 0.8	87.9
Path length [m]	43.6 ± 0.6	43.5
Avg. speed [m/s]	0.61 ± 0.02	0.50
Speed jerk RMS	22.3 ± 1.1	19.7
Position RMSE [m]	0.54	
Heading RMSE [rad]	0.31	
Discrete Fréchet Distance [m]	1.91	
Mean Leader–Follower Distance [m]	1.15	
Recovery time after vision loss [s]	0.2–0.5	

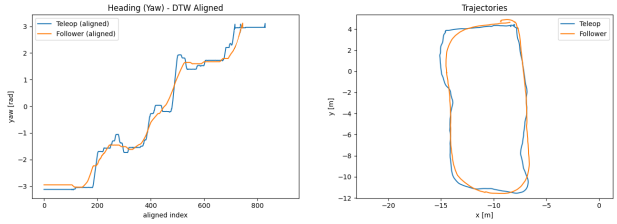

Fig. 11. Teleop vs. follower heading and trajectory comparison.

single-tag setup, azimuth accuracy is comparable to prior reports (approximately 2.1° outdoors and 4.5° indoors [40]); RSSI-based ranging was unreliable and excluded from closed-loop control. **Lessons learned.** Engineered priors (e.g., forearm color cues) and conservative control policies improve safety and yielding predictable recovery in field conditions, at the cost of reduced autonomy under vision-denied scenarios.

IX. CONCLUSION

This paper demonstrated a deployable quadruped robot follower system that enables hands-free, low-cognitive-load operation through a smartwatch interface. Experiments show that smartwatch control is practical and intuitive, while multimodal sensing improves robustness: vision provides metric following, and Bluetooth AoA supplies directional guidance when visual tracking is degraded. The multi-modal framework improved target re-identification in crowded scenarios, reducing tracking interruptions by (≈70%) and achieved reliable depth-based ranging (≈95%), though performance remains sensitive to lighting. Overall, integrating wearable control, a vision-based robot follower, and Bluetooth directional recovery cues highlights the potential of robust human-following robots for real-world deployment.

ACKNOWLEDGMENT

Support for this work was provided by the NIWC Pacific, the Naval Innovative Science and Engineering (NISE) program, and the DoD SMART Scholarship retention program.

REFERENCES

[1] M. Pollack, S. Engberg, J. T. Matthews, S. Thrun, L. Brown, D. Colbry, C. Orosz, B. Peintner, S. Ramakrishnan, J. Dunbar-jacob, C. Mccarthy, M. Montemerlo, J. Pineau, and N. Roy, "Pearl: A mobile robotic assistant for the elderly," in *Proceedings of AAAI '02 Workshop on Automation as Caregiver: the Role of Intelligent Technology in Elder Care*, July 2002.

[2] M. S. Shashank, P. Saikrishna, G. P. Acharya, S. Reddy, and P. Lavanya, "Design and development of human following autonomous airport baggage transportation system," in *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, 2022, pp. 1–6.

[3] R. Murphy, "Trial by fire [rescue robots]," *IEEE Robotics & Automation Magazine*, vol. 11, no. 3, pp. 50–61, 2004.

[4] Skydio Inc., "Skydio autonomy: Advanced flying ai," https://www.skydio.com/skydio-autonomy, 2023, accessed: 2025-06-06.

[5] A. Kostić, M. Schörghuber, M. Hofstätter, and M. Wallner, "Autonomous following wheelchair for passengers with reduced mobility at airports," in *2025 11th International Conference on Mechatronics and Robotics Engineering (ICMRE)*, 2025, pp. 82–89.

[6] Unitree Robotics, *Unitree Go1*, Unitree Robotics, 2021, accessed: 2025-09-11. [Online]. Available: https://www.unitree.com/go1

[7] ——, *Unitree Go2*, Unitree Robotics, 2023, accessed: 2025-09-11. [Online]. Available: https://www.unitree.com/go2

[8] H. Martins and R. Ventura, "Immersive 3-d teleoperation of a search and rescue robot using a head-mounted display," in *Proceedings of the 2009 IEEE Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE, 2009, pp. 1–6.

[9] F. C. Weigend, X. Liu, S. Sonawani, N. Kumar, V. Vasudevan, and H. Ben Amor, "iroco: Intuitive robot control from anywhere using a smartwatch," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 17 800–17 806.

[10] F. C. Weigend, N. Kumar, O. Aran, and H. Ben Amor, "Wearmocap: Multimodal pose tracking for ubiquitous robot control using a smartwatch," *Frontiers in Robotics and AI*, vol. 11, p. 1478016, 2025.

[11] F. C. Weigend, S. Sonawani, M. Drolet, and H. B. Amor, "Anytime, anywhere: Human arm pose from smartwatch data for ubiquitous robot control and teleoperation," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 3811–3818.

[12] T. Aziz and K. Insoo, "Enhancing indoor localization accuracy through multiple access point deployment," *Electronics*, vol. 13, no. 16, 2024. [Online]. Available: https://www.mdpi.com/2079-9292/13/16/3307

[13] P. Singh, M. Jain, L. Rawal, and P. Nagrath, "ibeacon-based indoor positioning systems for airports," *International Journal of Computer Applications*, vol. 179, no. 43, pp. 52–55, May 2018. [Online]. Available: https://ijcaonline.org/archives/volume179/number43/29375-2018917058/

[14] M. Cominelli, P. Patras, and F. Gringoli, "Dead on arrival: An empirical study of the bluetooth 5.1 positioning system," in *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, ser. WiNTECH '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 13–20. [Online]. Available: https://doi.org/10.1145/3349623.3355475

[15] B. V. Pradeep, E. S. Rahul, and R. R. Bhavani, "Follow me robot using bluetooth-based position estimation," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017, pp. 584–589.

[16] A. Satan and Z. Tóth, "Development of bluetooth based indoor positioning application," in *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)*, 01 2018, pp. 1–6.

[17] X. Zhang, Y. Chen, M. T. Hassan, and K. Suzuki, "Peer-to-peer ultra-wideband localization for hands-free control of a human-guided smart stroller," *Sensors*, vol. 24, no. 15, p. 4828, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/15/4828

[18] M. Ahmad, S. Alhady, S. Kaharuddin, and W. Othman, "Visual based sensor cart follower for wheelchair by using microcontroller," in *2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2015, pp. 123–128.

[19] R. Algabri and M.-T. Choi, "Deep-learning-based indoor human following of mobile robot using color feature," *Sensors*, vol. 20, no. 9, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/9/2699

[20] M. Gupta, S. Kumar, L. Behera, and V. K. Subramanian, "A novel vision-based tracking algorithm for a human-following mobile robot," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 7, pp. 1415–1427, 2017.

[21] M. K. Uddin, A. Bhuiyan, F. K. Bappee, M. M. Islam, and M. Hasan, "Person re-identification with rgb–d and rgb–ir sensors: A comprehensive survey," *Sensors*, vol. 23, no. 3, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/3/1504

[22] M. J. Islam, J. Hong, and J. Sattar, "Person-following by autonomous robots: A categorical overview," *The International Journal of Robotics Research*, vol. 38, no. 14, pp. 1581–1618, 2019. [Online]. Available: https://doi.org/10.1177/0278364919881683

[23] Z. Imani, H. Soltanizadeh, and A. Orouji, "Short-term person re-identification using rgb, depth and skeleton information of rgb-d sensors," *Iran Journal of Science and Technology, Transaction of Electrical Engineering*, vol. 44, pp. 669–681, 2020.

[24] G. Zidani, D. Djarah, A. Benmakhlouf, and L. KHETTACHE, "Optimizing pedestrian tracking for robust perception with yolov8 and deepsort," *Applied Computer Science*, vol. 20, pp. 72–84, 03 2024.

[25] W. Rahmaniar and A. Hernawan, "Real-time human detection using deep learning on embedded platforms: A review," *Journal of Robotics and Control (JRC)*, vol. 2, 11 2021.

[26] Boston Dynamics, "Spot® agile mobile robot," https://www.bostondynamics.com/products/spot, 2020, accessed: 2025-06-07.

[27] Intel Corporation, "Intel realsense depth camera d455," https://www.intelrealsense.com/depth-camera-d455/, 2020, accessed: 2025-06-06.

[28] Silicon Labs, "Bg22 bluetooth dual polarized antenna array pro kit (bg22-pk6022a)," https://www.silabs.com/development-tools/wireless/bluetooth/bgm22-pro-kit, 2020, accessed: 2025-06-06.

[29] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open-source robot operating system," in *ICRA Workshop on Open Source Software*, 2009, pp. 1–6.

[30] R. C. Smith, M. Self, and P. Cheeseman, "A Solution to the Simultaneous Localization and Map Building Problem," *IEEE Transactions on Robotics and Automation*, vol. 6, no. 5, pp. 589–599, 1990.

[31] J. Y. Raheema, M. R. Hess, R. C. Provost, M. Bilinski, and H. I. Christensen, "Autonomous exploration and mapping payload integrated on a quadruped robot," in *Proceedings of the International Symposium on Robotics Research (ISRR)*, dec 2024. [Online]. Available: https://www.cogrob.org/publication/raheema-24/

[32] E. Olson, "Apriltag: A robust and flexible visual fiducial system," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 3400–3407.

[33] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[34] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics yolov8," 2023, aGPL-3.0 License. [Online]. Available: https://github.com/ultralytics/ultralytics

[35] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Learning generalisable omni-scale representations for person re-identification," *TPAMI*, 2021. [Online]. Available: https://github.com/KaiyangZhou/deep-person-reid

[36] Google AI Edge, "Mediapipe," https://github.com/google-ai-edge/mediapipe, 2025, accessed: 2026-05-15.

[37] M. D. Hoang, S. S. Yun, and J. S. Choi, "The reliable recovery mechanism for person-following robot in case of missing target," in *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, 2017, pp. 800–803.

[38] C.-Y. Yang, H.-W. Huang, W. Chai, Z. Jiang, and J.-N. Hwang, "Samurai: Motion-aware memory for training-free visual object tracking with sam 2," *IEEE Transactions on Image Processing*, vol. 35, pp. 970–982, 2026.

[39] H. Alt and M. Godau, "Computing the fréchet distance between two polygonal curves," *International Journal of Computational Geometry & Applications*, vol. 5, no. 1-2, pp. 75–91, 1995.

[40] Silicon Laboratories Inc., "An1195: Antenna array design guidelines for direction finding," Silicon Laboratories Inc., Tech. Rep. AN1195, June 2022, revision 0.4. [Online]. Available: https://www.silabs.com/documents/public/application-notes/an1195-antenna-array-direction-finding.pdf