IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 January 2025, accepted 25 February 2025, date of publication 5 March 2025, date of current version 13 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3548134

SURVEY

Human Following and Guidance by Autonomous Mobile Robots: A Comprehensive Review

ANDREA EIRALE, (Member, IEEE), MAURO MARTINI, (Member, IEEE), AND MARCELLO CHIABERGE, (Member, IEEE)
Department of Electronics and Telecommunications (DET), Politecnico di Torino, 10129 Turin, Italy
Corresponding author: Andrea Eirale (andrea.eirale@polito.it)

ABSTRACT Collaborative and companion robots are at the forefront of technological innovation, transforming human-robot interaction to address a wide range of tasks and activities. This review provides a comprehensive examination of the current state of research on human following and guidance with autonomous mobile robots. Covering the evolution of research from the inception to the latest advancements in this area, we categorize existing literature based on various attributes including fields of application, technologies employed, and social acceptability. We critically analyze and compare state-of-the-art approaches in perception, tracking, planning, control, and human-robot interaction, highlighting their effectiveness and feasibility. We further classify studies based on application domains where person following and guiding tasks are particularly impactful, such as healthcare, personal assistance, logistics, and tour guiding. We identify persistent challenges and outline open problems, offering recommendations for future research directions. Our review aims to serve as a foundational reference for researchers and practitioners, fostering continued innovation and development in the deployment of autonomous robots for human following and guidance.

INDEX TERMS Autonomous robots, collaborative robotics, human–robot interaction, person following, robot guidance, survey.

I. INTRODUCTION

Autonomous service robots are emerging as the frontier technology to enhance human well-being. Despite advancements in robotic design and manufacturing revolutionized the field, intelligence remains the cornerstone for integrating autonomous robots into our daily lives. According to Artificial Intelligence experts, physical or embodied AI is a critical challenge for the next wave of innovation [1], [2]. As a result, robotics research is set to play a central role in global efforts to provide mobile and service robots with advanced capabilities that can meaningfully contribute to society.

Service robots are primarily designed to assist humans in social environments or demanding tasks. In healthcare, robots revolutionized patient care by assisting medical staff with the transportation of medical supplies [3] or by guiding patients in the medical environment [4]. These robots reduce physical strain and allow healthcare professionals to focus more on patient care. In logistics, robots streamline warehouse operations by following workers to carry heavy loads, optimizing workflow, and reducing the risk of workplace injuries [5]. Retail environments are also witnessing the integration of service robots, where they enhance customer assistance by guiding shoppers through stores, providing product information, and even carrying purchases [6], [7]. Moreover, in the realm of personal assistance, robotic platforms offer invaluable support to individuals with mobility challenges [8], following owners through their homes [9], and even assisting with daily tasks [10], [11].

An essential capability of intelligent assistive robots in these applications is to follow or accompany humans to specific destinations effectively. When following, the humans lead the task, and the robot pursues them. Typically, the robot trails behind the human, retracing their path.

The associate editor coordinating the review of this manuscript and approving it for publication was Francesco Mercaldo.

Alternatively, the robot may accompany the person side-by-side, which often results in more natural and effective Human-Robot Interaction (HRI) but introduces additional challenges. Although following and guiding constitute two different tasks, they share several similarities [12]. In both cases, the robot must maintain proximity to the human to keep track of them while ensuring a socially acceptable distance by adapting its velocity to the user's walking pace. Both following and guiding tasks necessitate sophisticated sensor arrays and advanced algorithms for environmental perception, human detection, and navigation. Robots must be able to distinguish between animate and inanimate objects, recognize human intentions, and make split-second decisions to avoid obstacles and ensure a smooth interaction. Contextually, approaching the extensive literature on autonomous navigation, perception, control, and human-robot interaction can be dramatically hard and time-consuming for a novice researcher. Therefore, we believe that a focused survey highlighting the latest advancements and emerging trends in human-following and guidance by autonomous robots is an essential and valuable resource for the service robotics community.

This article builds on and extends previous surveys in the field [13], [14]. The most recent survey, conducted by Islam et al. in 2019 [14], categorized studies based on operational medium–ground, underwater, or aerial–and further classified them by factors such as sensor type, interaction mode, number of robots and humans involved, and autonomy level. However, we believe many of these categorization factors are often more closely tied to the specific application domains of the robotic platform. This observation led us to adopt a different approach in our survey. We organized the categorization around core methods and enabling technologies that are fundamental across all domains and critical for effectively implementing person following and guiding robots. Moreover, we expanded the scope of our survey by integrating robot guiding tasks alongside person following, reflecting a more comprehensive understanding of human-centered navigation. In the second section of the paper, we also re-organized studies based on their most relevant application fields, discussing the unique challenges and advancements within each of them.

Therefore, this paper provides a detailed exploration of person following and guiding robots, examining their broad applications and potential to reshape human-robot interaction, ultimately influencing the way we live and work. In particular, the contributions of this paper are manifold:

- We systematically review the scientific literature on human following and guidance by mobile, autonomous robots, spanning from early research to the latest advancements. Our review offers an in-depth comparison and discussion of state-of-the-art approaches across key areas essential to the functioning of person following robots. These include perception (how robots sense and interpret their environment), tracking (how robots continuously monitor and follow the target individual), planning (how robots determine optimal paths), control (how robots execute physical movement), and interaction (how robots communicate and engage with humans). Our comparison focuses on evaluating the effectiveness and feasibility of different approaches, providing insights into their practical applications and potential limitations.
- We categorize these works according to their application domains, highlighting the primary usage and specific functions of person following and guiding robots in various fields such as healthcare, personal assistance, logistics, and tour guiding. This categorization helps clarify each field's specific challenges and requirements, allowing for a more tailored understanding of the technologies employed.
- We present and analyze the primary evaluation methods used in literature to assess the performance of these systems. Each technology area–perception, tracking, planning, control, and interaction–has its own set of performance metrics and experimental setups. Then, we critically evaluate the solutions proposed in literature, focusing on their feasibility, practicality, outcomes, and costs.
- Finally, we identify and highlight the open challenges that persist in the field of person following and guiding robots. These challenges include improving the system's adaptability to different contexts, such as when the robot needs to follow a group of people, take the lead in guiding an individual, and maintain high speeds during pursuit. Based on our review and analysis, we provide a set of recommendations for future research directions. These suggestions aim to guide researchers in overcoming current limitations, addressing technical gaps, and pushing the boundaries of what person following robots can achieve.

A. REVIEW METHOD

We conducted a systematic review of the literature on human following and guiding robots to provide a comprehensive synthesis of the state-of-the-art in this field. This survey is intended to assist researchers in analyzing and organizing previous work on these topics while also identifying key trends, challenges, and advancements in robot deployment for following and guiding tasks. We selected only studies that specifically addressed human following and guiding applications tackling meaningful technical aspects such as perception, tracking, navigation, and interaction.

Our search was conducted using various search engines, such as Google Scholar, Scopus, IEEE, Science Direct, Sage Journals, and Springer. We used the following keywords: "Robot person following", "Robot human following", "Robot guide", "Robot guidance", and "Robot guiding". We limited the scope to peer-reviewed conference papers, journal articles, and book chapters, all published in English. No restrictions were placed on the starting publication date, as we aimed to cover the entire breadth of available literature on these topics. The initial search was carried out on 28 May

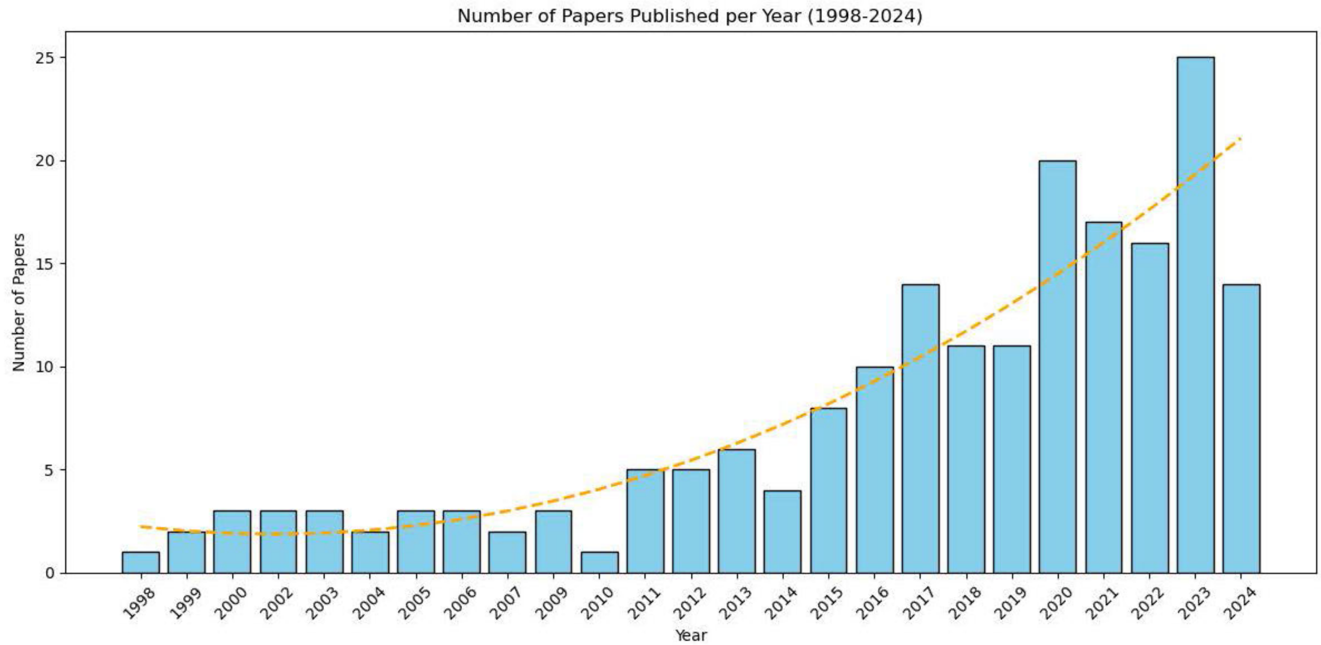


FIGURE 1. The yearly distribution of published papers in the field of human following and guidance with autonomous mobile robots.

2024, and a follow-up search was conducted on 08 January 2025.

In total, 482 works were retrieved and organized into a unique database, with metadata such as title, authors, publication date, venue, abstract, and citation count. After removing 16 duplicate entries, 466 unique works remained for review. We carefully evaluated each work, maintaining only those that explicitly focused on robots designed for person following and guiding tasks. Studies that primarily addressed perception, tracking, control, and/or human-machine interaction in a different context were excluded to preserve the relevance of the survey. Finally, a further screening was conducted based on the number of citations for each year. Older papers were only retained if their contribution remains relevant and holds significance in the context of more recent advancements in the field. After this rigorous selection process, 192 works were retained and form the core of the analysis presented in this paper. These studies represent the most relevant contributions to the field of human following and guidance by autonomous mobile robots and provide a foundation for future research. Figure 1 shows the annual distribution of these selected papers, demonstrating the increasing interest of research in the field. Moreover, the high rate of publication in the last two years reflects the real state of autonomous robot development: many unsolved challenges and practical tasks still have to be solved to fully launch the service robotics market.

II. METHODOLOGIES

Autonomous following and guiding tasks require a comprehensive suite of sensors and sophisticated algorithms to perform essential submodules: perception, detection and tracking, path planning, control, and human interaction. Perception enables the robot to understand its environment, detect obstacles, and consistently recognize and track the target person. The navigation system enables the robot to plan an efficient and safe path while maintaining social norms during movement. Planning and control modules work together to ensure the robot takes an optimal, obstacle-free route toward the goal while adjusting for real-time conditions and human behavior. Interaction is another vital component, allowing the robot to receive commands, communicate effectively, and provide socially appropriate responses.

In this section, we categorize and examine state-of-the-art approaches in these critical areas, central to the functionality of person following and guiding robots: perception, tracking, planning, control, and interaction. Figure 2 provides an overview of the methods discussed in this chapter.

A. PERCEPTION

The perception module plays a crucial role in autonomous systems. It is responsible for both detecting obstacles around the robot and identifying the target person to be tracked. The module must integrate various sensing techniques to create a cohesive and accurate environment representation, ensuring smooth navigation and interaction with humans. Typically, obstacle detection relies on Time-of-flight (TOF) sensors, which provide real-time distance measurements by calculating the time it takes for emitted signals to bounce back from surrounding objects. Among TOF sensors, Laser Range Finder (LRF) and LiDAR systems are the most commonly used in robotic applications [16], [19], [71], [73], [74], [75], [76], [77], [78]. These sensors offer high precision

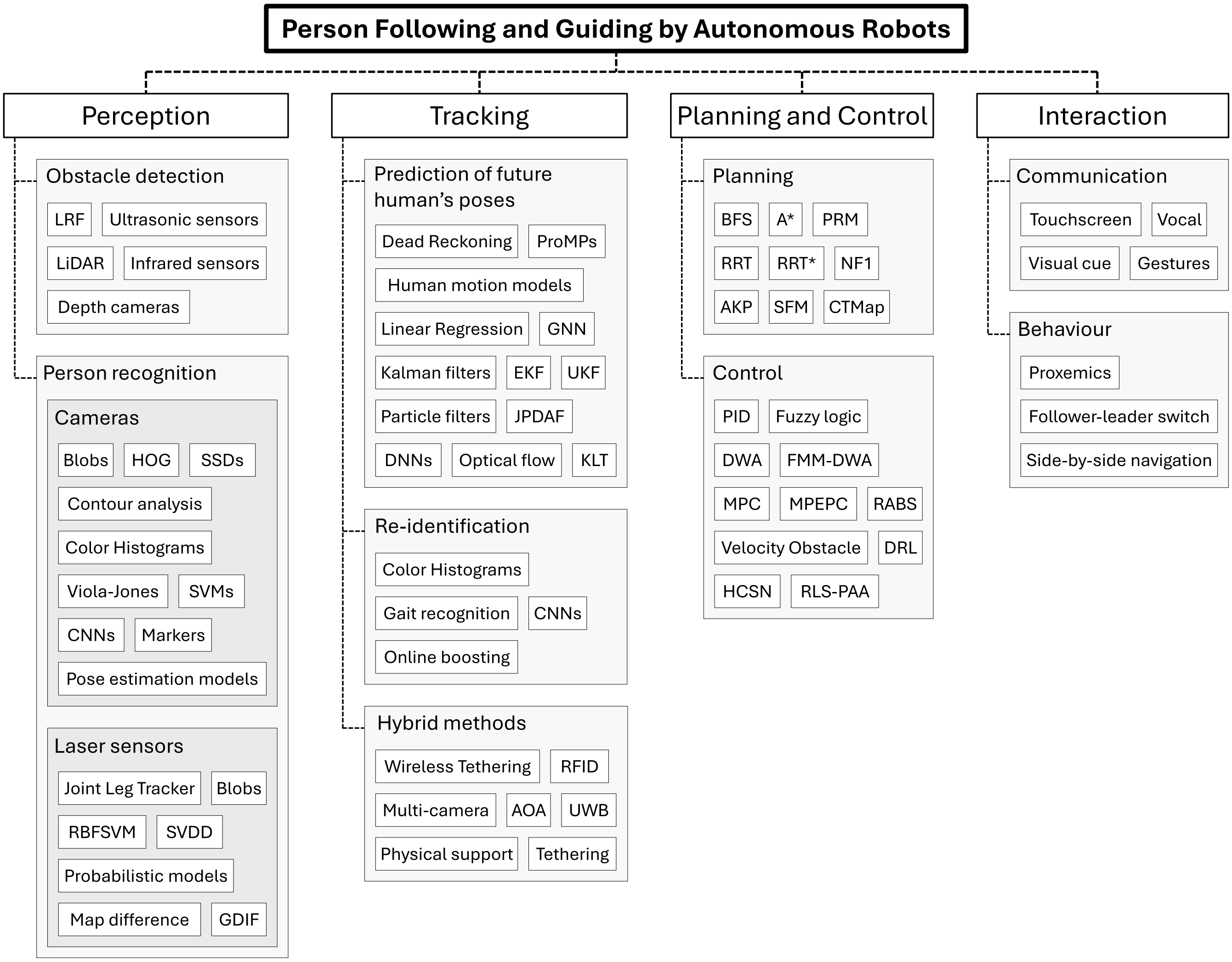


FIGURE 2. A categorization of the state-of-the-art methods and approaches for person following and guiding autonomous robots, divided in four main key areas: perception, tracking, planning and control, and interaction. For acronyms, refer to Section VI.

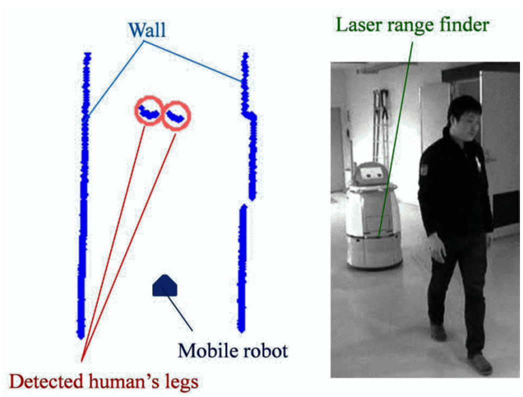


FIGURE 3. An example of 2D Laser Range Finder detection, including obstacles such as walls and human legs. [3].

and wide field-of-view capabilities, making them effective for both indoor and outdoor navigation, even in cluttered environments. Figure 3 represents a typical LRF detection.

Ultrasonic sensors are another popular choice for obstacle detection, particularly in environments where cost and simplicity are critical factors. Although ultrasonic sensors provide lower resolution compared to LRF or LiDAR, they are effective at short distances and can detect objects in various lighting conditions [4], [36], [41], [79], [80], [81]. They are often combined with other sensors to enhance the overall reliability of the perception system. Similarly, infrared sensors, though less common, are employed in some applications where low-cost and simple obstacle detection are sufficient [82]. These sensors are particularly suited for detecting objects with strong thermal signatures or reflecting properties. In addition to traditional TOF sensors, some works leverage depth cameras, which generate detailed 3D depth maps by analyzing stereo images or structured light patterns. Depth cameras provide rich spatial information that enables the detection of obstacles at varying heights, making them particularly useful for complex environments such as urban areas or indoor spaces with multiple levels of objects [22],

TABLE 1. Perception and person recognition methods.

Obstacle detection sensors	Person detection sensors	Person detection data	Body part recognized	Features	Detection Method	References
LRF	LRF	Laser scan	Legs	Laser scan blobs	Geometric based filtering	[15]–[24]
LRF	LRF	Laser scan	shoulder	Shoulder laser model	Geometric based filtering	[25], [26]
LRF	LRF	Laser scan	Legs	legs shape's attributes	SVDD classification algorithm	[27]–[30]
-	Color camera	Color image	Shirt	Blobs	Color histogram	[31]–[33]
-	Color camera	Color image	Face	Blobs	Color based	[34]–[36]
-	Color camera	Color image	Whole body	HOG features	SVM classifier	[37]
-	Color camera	Color image	Marker	Color pattern	Color based	[38]–[40]
Ultrasonic array	Color camera	Color image	Marker	Color pattern	Color based	[4], [41]–[44]
-	Color camera	Monochrome image	Markers (Apriltag/Aruco)	Specific marker pattern	Marker detector	[5], [45]
RGBD camera	RGBD camera	Color depth image	Whole body	YOLO latent features	YOLO	[46]–[49]
LRF/LiDAR, RGBD camera	RGBD camera	Color depth image	Whole body	YOLO latent features	YOLO	[50]–[60]
-	Color camera	Color image	Whole body	Neural Network latent features	CNN classifier	[61], [62]
Ultrasonic array	RGBD camera	Color image	Whole body	Neural Network latent features	CNN classifier	[63]
-	Color camera	Color image	Whole body	OpenPose skeleton	OpenPose	[64]–[66]
LiDAR	RGBD camera	Color image	Whole body	OpenPose skeleton	OpenPose	[6], [9], [10], [67]
-	Color camera	Color image	Face	Haar features	AdaBoost classifier	[68], [69]
LRF	Color camera	Color image	Face	Haar features	AdaBoost classifier	[70], [71]
LRF	LRF, RGBD camera	Laser scan, Color image	Legs, Whole body	Laser scan blobs, Neural Network latent features	Geometric based filtering, CNN classifier	[72]

[83]. Depth cameras can also assist in distinguishing between flat surfaces and protruding obstacles, which is essential for safe navigation.

While obstacle detection is well-understood and relatively straightforward, human detection poses a more significant challenge due to the complexity and variability of human appearances and behaviors. Unlike static obstacles, humans are dynamic entities that change their position, orientation, and even appearance over time. Therefore, detecting people requires sophisticated feature extraction and classification algorithms. The features used for detection depend on the sensor type and can range from general characteristics, like silhouettes or movement patterns, to more specific features, like facial landmarks, body joints, or even clothing colors. In the following subsections, we will explore the different methods used in literature for person perception, categorizing them into the two most common sensor types: camera-based and laser-based systems. Table 1 gathers the main perception and person recognition methods discussed.

1) CAMERA-BASED HUMAN DETECTION

Early approaches in visual person detection commonly used blob features, clusters of similar pixels in an image identified by properties like color, texture, and motion [84]. Blob

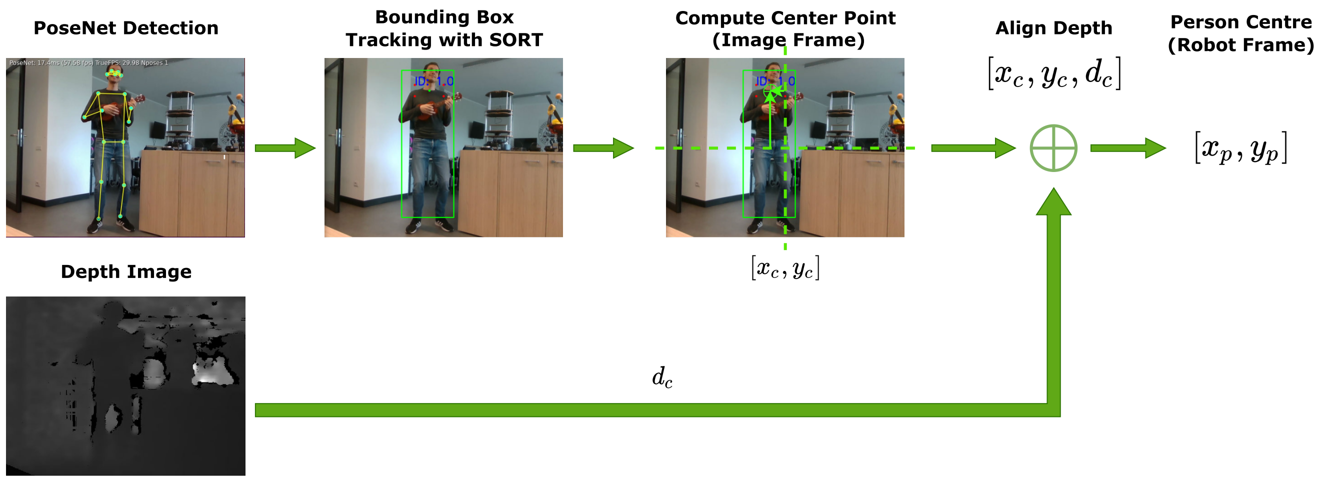


FIGURE 4. Skeletal outputs provide more accurate localization, enabling better alignment between color and depth images, ultimately resulting in more precise tracking of the user in 3D space [9].

features reduce image complexity to manageable parameters, facilitating object recognition in noisy environments [22], [32], [85]. Blob features are frequently combined with color histograms, which represent the distribution of colors in an image. By comparing the color histogram of a detected region with that of a known target, a robot can maintain its focus on the same person, even if other visual features are ambiguous or changing. High histogram values indicate that the color value strongly corresponds to the tracked object, providing a reliable cue for maintaining focus on the same person or object over time [31]. Despite their simplicity, blob features and color histograms are still widely used nowadays due to their computational efficiency and ease of implementation, making them a popular choice in real-time applications [76], [86], such as person re-identification [6], [87], marker detection [39], and object tracking [33]. Another popular color-based person detection approach involves thermal cameras [88], [89], [90], [91]. Thermal cameras detect the presence of humans and pinpoint their location within the environment by capturing body heat. This approach is particularly useful in scenarios where standard RGB-based detection may be unreliable, such as in smoke-filled, dark, or crowded environments [92]. Contour features, which capture object boundaries, were also widely used. While computationally demanding, contour-based methods yield high accuracy in dynamic settings when combined with blobs or other visual features [31]. As visual recognition advanced, contour features evolved into the most advanced Histogram of Oriented Gradients (HOG) features. Due to their effectiveness in capturing edge and gradient information [93], HOG features are widely adopted in various applications, including pedestrian detection and security surveillance [37], [49].

Early face detection systems often relied on a combination of blob features and color-based approaches, particularly those focusing on skin color. These methods provided a straightforward yet effective means of detecting faces by identifying regions in an image that matched typical skin tones [34], [35], [36]. However, early systems were susceptible to variations in lighting, background, and the diverse range of human skin tones, which limited their robustness and applicability in real-world scenarios. Despite these limitations, such foundational techniques played a critical role in the development of more advanced face detection algorithms. One of the most significant advancements was the introduction of the Viola-Jones object detection framework [94], [95]. Developed by Paul Viola and Michael Jones, this algorithm was revolutionary for its time, thanks to its ability to perform real-time face detection, a feat unattainable with earlier methods. The Viola-Jones algorithm leverages Haar-like features, simple rectangular features that capture differences in pixel intensity. Then, it employs a cascade of AdaBoost classifiers that progressively narrow down candidate regions in an image. Unlike color-based methods, the Viola-Jones algorithm does not depend on color information, making it more adaptable to varying lighting conditions and diverse environments [68], [69], [71].

In addition to face detection, visual recognition of markers applied to a person's clothing is another technique commonly used in literature. This method involves attaching visual markers—often distinct color patterns or shapes—to the target person's clothing, which can be easily identified by monocular cameras [4], [38], [39], [41], [42]. By knowing the dimensions of these markers, it is possible to estimate the distance and position of the person without additional depth sensors. Several types of visual markers have gained popularity in this context, each suited to different applications. For example, AprilTags and ArUco markers are widely used for their robustness and ease of detection [5], [45]. These markers consist of binary patterns that are highly distinctive, making them easily recognizable even in cluttered environments. Infrared (IR) markers offer the advantage of being detectable in low-light or night-time conditions, as they emit infrared light that can be picked up by specialized cameras [43]. Other generic visual markers are often designed for specific use cases, such as indoor navigation, where simplicity and reliability are paramount [40], [44]. Visual marker recognition is particularly valuable in scenarios where other forms of detection may struggle, such as in environments with poor lighting, or when the subject's appearance may change frequently, such as in sporting events. The use of markers provides a level of robustness and simplicity that is essential in many practical applications, ensuring reliable person recognition and tracking even under challenging conditions.

As machine learning techniques evolved, methods like Support Vector Machine (SVM) became widely adopted for classifying people based on features like blobs, Histograms of Oriented Gradients, and color histograms [37], [96]. SVMs offer robust generalization, especially when combined with kernel tricks, enabling accurate classification in high-dimensional feature spaces. In recent years, deep learning models have revolutionized object detection by offering end-to-end solutions capable of directly extracting human features from Red-Green-Blue (RGB) images and providing high-accuracy bounding boxes. Single Shot Detector (SSD)s [97] and models like MobileNet-SSD [98] have become highly popular due to their balance between speed and precision, making them well-suited for real-time applications on resource-constrained devices [72], [76], [99], [100], [101]. Another prominent multibox object detection model is You Only Look Once (YOLO) [102], which gained widespread usage because of its efficiency in detecting objects, including people, in a single pass through the network [46], [47], [48], [50], [51], [52], [54], [55], [56], [57], [58], [59], [60]. The versatility of YOLO in detecting multiple object classes simultaneously and its adaptability to various deployment scenarios has made it a go-to solution in numerous robotics and surveillance applications [49]. Ongoing research focuses on optimizing these models for better performance on edge devices [53], [103], ensuring high processing speeds without sacrificing accuracy. Key optimization techniques include model quantization, which reduces the precision of weights and activations, and pruning, which removes redundant neurons or connections. Other powerful alternatives in person detection are segmentation networks like Mask R-CNN [104], which offer pixel-level precision by generating instance-aware segmentation masks around detected objects [105]. Mask R-CNN builds upon the Faster R-CNN [106] architecture, extending it by adding a parallel branch for object mask prediction alongside the bounding box and class prediction.

In addition to pre-trained object detection models, custom solutions based on CNNs have been introduced for specific person detection tasks [61], [62], [63], [107]. These custom CNN models are typically fine-tuned for particular environments or applications, allowing for improved performance in domain-specific scenarios, such as in challenging lighting conditions or crowded scenes. Custom architectures also provide flexibility in integrating additional features like multi-view consistency, temporal tracking, and context-aware detection, making them powerful tools for specialized tasks. For applications where understanding human movement and interaction is critical, pose estimation models are increasingly preferred over traditional bounding box methods [6], [64], [65], [66], [67], [73], [108], [109], [110]. These models predict the skeletal structure of a person by estimating key joint positions, such as elbows, shoulders, knees, and feet, as can be seen in Figure 5. Notable models in this category include PoseNet [111], OpenPose [112], MoveNet [113], and the Kinect's Microsoft SDK [114]. OpenPose excels at capturing detailed human body poses, including multiple people in a scene, making it particularly useful in crowded environments where multiple individuals must be distinguished. Similarly, MoveNet is optimized for fast and accurate pose estimation on edge devices, offering a compact and efficient solution for mobile robots and wearables that need to track user movement in real-time. The Kinect SDK, though originally designed for gaming, remains a versatile tool in robotics and research due to its robust skeletal tracking capabilities and wide adoption. Specifically, these models are often preferred over bounding boxes due to their greater accuracy [9], [10], As can be seen in Figure 4, key joints robustly fall over the person being tracked. Conversely, not all pixels contained within the bounding box belong to the tracked subject. This could lead to the robot following a point in the background of the scene instead of the actual person of interest.

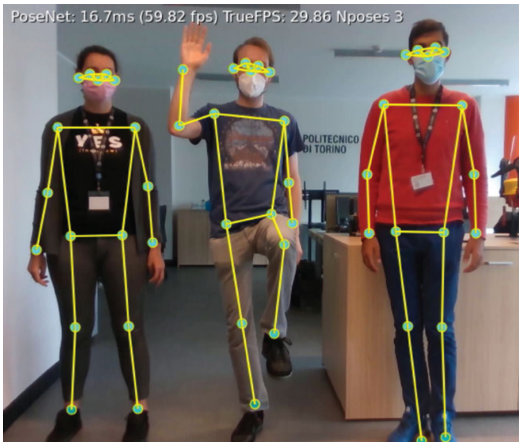


FIGURE 5. Example of skeletal structure of multiple individuals predicted by PoseNet.

2) LASER-BASED HUMAN DETECTION

Laser sensors, including LRF and LiDAR, are integral tools in person recognition and localization tasks. Depending on the specific application, different body parts may be targeted for detection. For instance, shoulders are particularly suitable in side-by-side navigation, as highlighted in works by Kobayashi and others [25], [26]. Accurate shoulder positioning facilitates smoother and safer collaborative tasks for

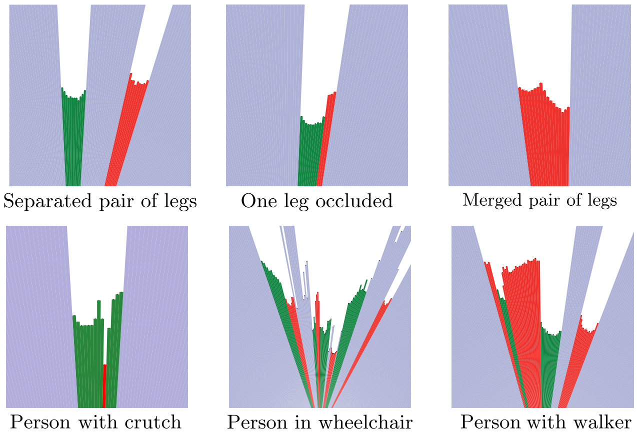


FIGURE 6. Demonstration of Generic Distance-Invariant Features to model a variety of leg configurations for person detection [116].

applications involving close human-robot interaction. However, the majority of research in this domain has concentrated on leg detection, given the distinctiveness and consistency of leg motion during walking. Leg detection methods predominantly rely on geometric approaches. In these techniques, laser scan data points are first segmented and clustered into blobs based on spatial proximity. Then, they are classified using probabilistic models that retain clusters corresponding to human dimensions while discarding those that do not meet predefined criteria [15], [19], [22], [96]. The geometric approach is particularly effective in real-time applications due to its relatively low computational complexity. For instance, Miura et al. [3], [115] employed Radial Basis Function Support Vector Machine (RBFSVM) to classify laser data clusters. This method utilizes a kernel function to map input features into higher-dimensional spaces, enabling more accurate separation between human and non-human clusters. On the other hand, random forest classifiers have also been widely adopted for leg detection [77]. These classifiers offer robustness by integrating decision trees that collectively assess the likelihood of a cluster belonging to a human leg based on multiple feature inputs. In addition to geometric methods, specific leg patterns can be identified directly from the laser scan data. Hata et al. [17] proposed techniques that leverage the unique geometric patterns of legs within laser scans, such as the arc-shaped contour formed when a person is walking, to improve detection accuracy. Chen et al. [96] designed a double CNN model to select pedestrian clusters from laser features. The first network processes 2D laser information, while the second one processes voxelized features to obtain 3D information. The two outputs are then concatenated to improve detection accuracy while ensuring real-time performance. These approaches enhance robustness in environments where partial occlusion and overlapping clusters often lead to ambiguities.

Generic Distance-Invariant Features (GDIF) represent a more sophisticated method for recognizing human legs. As discussed in [116], GDIF allows the system to model a variety of leg configurations. Figure 6 illustrates some of these configurations, including scenarios where legs are separated, merged, or when one leg is occluded. GDIF can be integrated with specific detectors to enhance recognition capabilities from laser scan data [87]. This feature set is particularly useful in dynamic and cluttered environments, where the relative positions and orientations of legs can vary significantly. Another advanced system is the Joint Leg Tracker developed in [23], which applies clustering techniques to laser detection data while maintaining confidence levels to minimize false positives. The system was further refined in recent works [18], [24] by incorporating probabilistic data association mechanisms that improve tracking stability even in crowded environments. Support Vector Data Description (SVDD) is another classifier frequently employed in leg detection tasks. Unlike traditional SVM, which separates classes by maximizing the margin, SVDD constructs a hypersphere around the target class, effectively encompassing leg clusters while rejecting outliers. This approach is particularly well-suited for identifying human legs in noisy datasets, where standard classification techniques might falter. The first implementations of SVDD for leg detection were based on simple 2D features derived from laser scan data [27], [28], [30]. However, recent advancements by Cha et al. [29] extend this approach by mapping 2D LiDAR data into a 3D feature space. By constructing a dataset composed of 3D feature vectors of leg segments, they trained an SVDD algorithm to reconstruct 3D leg structures from the 2D scan data, enabling more robust recognition in complex environments. Another innovative approach in leg detection is map differencing, where human presence is inferred by detecting significant deviations in laser scan data from a pre-learned map of the static environment. This technique, initially explored by Pollack and Pineau [117], [118], remains highly relevant for applications in environments where static and dynamic elements need to be clearly distinguished, such as in warehouses or crowded urban areas. By comparing real-time laser scans with a pre-recorded map, deviations can be detected and classified as human presence, bypassing the need for more complex feature extraction and classification steps.

In modern systems, laser-based leg detection is increasingly integrated with multi-modal sensor setups, including Red-Green-Blue-Depth (RGBD) cameras, stereo vision, and Inertial Measurement Unit (IMU) [20], [21], [92]. These configurations enable the fusion of 2D and 3D data, providing richer context and reducing false positives. For instance, in [119], [120], authors continuously detect the user using a laser-based leg detection algorithm alongside a monocamera visual detector. Then, they construct a tracker that exploits an abstract metric space to fuse LRF and monocular camera data in a single detection output. A different approach is implemented in [72], where a hierarchical methodology is introduced. A laser-based leg detector is used to track people under normal conditions. In challenging scenarios, such as obstructions or sudden direction changes, a visual detector compensates for the limitations of the laser-based system.

Other non-conventional multi-sensor perception systems relying on audio devices and thermal cameras are often employed in unstructured and dynamic environments, where counting solely on laser data may lead to inconsistencies. For example, authors in [88] employ a particle filter to fuse RGB, laser, and thermal data to overcome detection limitations due to poor environmental lightning. Differently, in [70], audio information from two directional microphones is integrated with laser and visual information to better detect individuals requiring the robot's attention. Although multi-modal detection methods increase the robustness and accuracy of person detection, choosing a good policy to fuse the information coming from different sensors is not always trivial. In [121], Siva et al. introduced an innovative learning-based perception framework designed to optimally fuse multisensory input data. Their approach adapts to both short-term and long-term environmental changes by dynamically estimating the contribution of each feature to represent the environmental context. This estimation informs a sensor calibration process, assigning greater importance to observations derived from the most representative features.

B. TRACKING

Once a person is detected and localized within the environment, continuous tracking becomes essential for practical guidance and interaction. Tracking ensures that the system maintains an accurate, real-time understanding of the person's movements and position, which is crucial in dynamic environments where the individual may frequently change direction or speed. From a theoretical standpoint, tracking can be conceptualized as a sequence of continuous detections of the followed person [33]. However, practical complications arise when detection is intermittently lost due to obstructions, inefficiencies in the perception system, or environmental factors. In such instances, the risk of irretrievably losing track of the person increases, severely compromising the system's ability to guide and follow effectively. This challenge becomes even more pronounced in environments with multiple individuals, where the system must accurately identify and continue tracking the specific person of interest without confusion. To address these challenges, a reliable tracking system must not only handle intermittent detection failures but also minimize the reliance on re-identification modules. By integrating sophisticated algorithms that account for occlusions and dynamically changing conditions, the system can predict the future positions of the human companion, ensuring continuous and accurate tracking even under challenging circumstances.

Prediction of future human poses can be simply accomplished with dead reckoning, assuming constant velocity and direction over time [70], [122], a strategy that works reasonably well in controlled environments or for short time horizons. However, real-world scenarios involve more complex and non-linear human movement patterns, requiring more sophisticated prediction models. Many advanced approaches use human walking models derived from datasets containing real-life human trajectories [27], [63], [123], [124]. These models go beyond basic assumptions by learning typical patterns and behaviors in human motion, as illustrated in Figure 7. Starting from a history of human states, the models can predict future human positions over a defined time horizon. These states usually contain information about positions, velocities, and directions of the tracked person. However, sometimes more detailed features are included, such as bounding boxes [47], [125] and body joint coordinates [56], enabling more accurate predictions even in dynamic and cluttered environments. This data-driven approach is especially valuable in scenarios that demand human-aware navigation, where the robot must anticipate human movement and adapt its behavior accordingly. In these cases, the robot may use side-by-side [126], [127], [128], [129] and collaborative [130] navigation, which is more natural and socially acceptable in many contexts. Here, the challenge lies in predicting not just immediate movements but also longer-term destinations. Predicting such destinations involves a higher level of contextual understanding, where the robot must interpret cues from the human's actions, surroundings, and interaction history. Methods leveraging probabilistic reasoning, deep learning, and trajectory forecasting are often employed to infer potential goals or pathways. For instance, some approaches utilize Markov models or recurrent neural networks to analyze historical trajectory data and predict future movements, even in complex environments with multiple possible destinations [40], [127], [131], [132], [133], [134], [135].

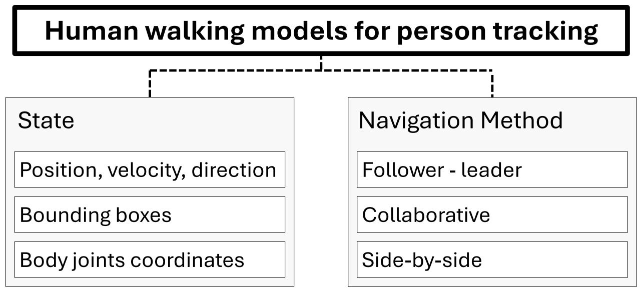


FIGURE 7. Categorization of human walking models. These models leverage different state information to predict future human motion for various types of navigation.

However, predictive models alone are often insufficient for efficient tracking. To achieve more robust and accurate tracking, they are usually integrated with statistical estimation algorithms that can handle uncertainties and adapt to dynamic environments. Among these algorithms, linear regressors [50], [99] and Probabilistic Movement Primitives (ProMPs) [136] are commonly used. Nevertheless, Kalman filter-based methods remain the most widespread due to their versatility and effectiveness in sensor fusion and noise reduction [5], [46], [60], [65], [86], [137], [138]. The Kalman filter's strength lies in its ability to predict and correct estimates by recursively updating the state of the tracked person using incoming sensor data. Classic variants like the Extended Kalman Filter (EKF) [24], [125] and the Unscented Kalman Filter (UKF) [52], [64], [96] extend the basic Kalman filter to handle the non-linear dynamics often present in human movement. These filters work well in scenarios where linear models are insufficient, such as when tracking irregular paths or rapidly changing velocities. The key advantage of these filters is their capacity for sensor fusion, which allows them to integrate data from multiple sources, such as vision, ultrasound, and inertial sensors, to produce more accurate estimates [77], [139]. Despite their effectiveness, Kalman filter-based solutions often encounter the data association problem. This issue arises when the filter must determine how to match new sensor measurements with existing tracks from previous time steps, particularly in crowded environments with multiple individuals. To address this problem, integration with techniques like Global Nearest Neighbor (GNN) [23], [51] or the Joint Probabilistic Data Association Filter (JPDAF) [28] is common. GNN offers a straightforward approach by selecting the most likely association based on proximity. In contrast, JPDAF provides a more probabilistic approach, considering multiple potential matches simultaneously and weighting them according to likelihood. An alternative to the Kalman filter is the particle filter, which is particularly suited for non-linear and non-Gaussian systems [71], [117], [118]. Unlike the Kalman filter, which relies on analytical equations, the particle filter employs a set of particles to represent the probability distribution of the tracked person's state. Each particle is a possible hypothesis of the current state, and the filter updates these particles based on sensor inputs and a motion model. The particle filter's strength lies in its flexibility. It can represent complex distributions and handle multi-modal scenarios where multiple hypotheses about the tracked person's location exist. However, this flexibility comes at the cost of higher computational demand, as the filter must simulate the evolution of many particles over time [88]. Additionally, particle filters are frequently applied in multi-person tracking scenarios, where they effectively manage the complexity of tracking multiple individuals in close proximity [15], [140]. In some cases, hybrid systems combine the strengths of both Kalman and particle filters. For example, in [37], each particle of a particle filter is processed using a simple Kalman filter. This combination allows the system to benefit from the Kalman filter's efficient estimation and correction capabilities while leveraging the particle filter's robustness in handling non-linearities and uncertainties.

Efficient artificial Deep Neural Network (DNN) have increasingly been adopted for tracking, mirroring their success in detection tasks. These neural networks offer robust solutions by learning complex motion patterns and adapting to non-linear dynamics. For example, in [144], a backpropagation neural network predicts future human positions based on past tracks obtained via a LRF. Once the neural network makes predictions, a cubic spline generates a

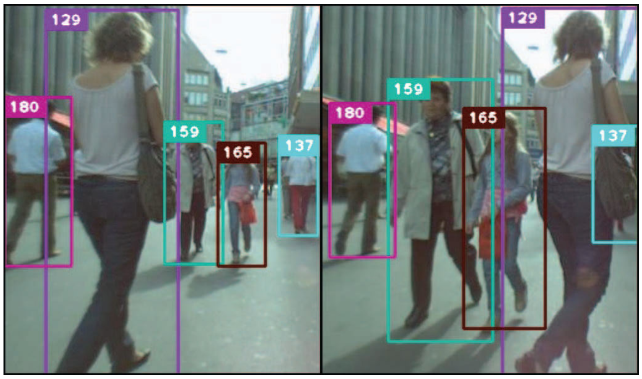


FIGURE 8. Visual demonstration of DeepSORT [141] tracking algorithm in a common crowded situation with frequent occlusion. The algorithm assigns a unique ID to each detected person and consistently maintains this ID throughout the entire tracking process.

smooth future path. This approach combines the adaptability of neural networks with traditional trajectory generation techniques, yielding a more accurate and reliable tracking solution. In more advanced systems, deep learning methods have expanded to predict not only positions but also complex human poses [110]. For instance, in [145], a non-autoregressive transformer model is used for simultaneous human trajectory and pose prediction in a following-ahead task. The system predicts the next positions of the tracked person and their future poses, including the 3D coordinates of all joints relative to a fixed hip joint. This holistic approach enables the robot to anticipate full-body movements, crucial for applications like assistive robots that must navigate near people.

While many tracking solutions in the literature focus on tracking a person's movement in 3D space, visual tracking methods offer the flexibility to track individuals within the 2D image space. The main advantage here lies in tracking individuals directly from visual data, such as bounding boxes derived from image detection, instead of relying on precise 3D-space localization. One prominent example is the Simple Online and Realtime Tracking (SORT) algorithm [146]. In SORT, bounding boxes generated by visual detection are used as inputs for the Kalman filter. This filter predicts and updates the object's position across frames, ensuring reliable tracking even when temporary occlusions occur [9], [147]. SORT was later enhanced to DeepSORT [141], which significantly improves tracking accuracy by using a deep association metric to match detections across frames, making it robust to appearance changes, occlusions, and similar-looking objects in crowded environments [59]. Figure 8 presents a visual demonstration of the DeepSORT algorithm, closely resembling the typical output of SORT. In this representation, the algorithm assigns a unique ID to each detected individual and reliably maintains these IDs throughout the entire tracking sequence. Optical flow-based methods offer another viable solution for visual tracking. These techniques, including those presented in [88] and [148], analyze the motion between successive video

TABLE 2. Person re-identification methods.

Person detection sensors	Person detection data	Body part recognized	Features	Re-identification Method	References
Color camera	Color image	Whole/upper body	HOG/color features	Color histogram	[6], [76], [87], [99]
RGBD camera	RGBD image	Legs	Skeletal joints and gait features	Gait recognition	[108], [115]
RGB camera	RGB image	Whole body	Neural network latent features	CNN Classifier	[50], [64], [142]
RGB camera	RGB image	Whole body	Neural network latent features	Attention-based Classifier	[143]
RGB camera	RGB image	Whole body	Convolutional Channel Feature	Online Boosting Classifier	[65]

frames to estimate the movement of objects. By tracking the pixel displacement, optical flow methods can estimate motion vectors and maintain tracking consistency, even when direct detection might be challenging due to occlusions or lighting conditions. Another well-established technique for visual tracking is the Kanade-Lucas-Tomasi (KLT) feature tracker, which extracts and tracks distinctive image features, such as corners or edges, across consecutive frames. The KLT algorithm leverages the assumption that small image regions containing these features exhibit consistent motion across frames. By comparing these small regions, KLT provides a reliable way to track moving objects over time, even in noisy environments [20], [68]. This method is particularly useful for applications requiring detailed motion information, such as gesture recognition or fine-grained activity tracking.

One of the most significant challenges in person tracking is handling situations where the tracked individual becomes partially or fully occluded. Occlusions can arise from various factors: the person may move behind an obstacle or turn around a corner, the robot may take a different trajectory while following, or another dynamic obstacle may temporarily block the line of sight between the robot and the person. In these situations, the robot needs to maintain a probabilistic belief about the person's location in its surroundings, even when direct visual information is unavailable due to total occlusion [47], [69]. In cases of partial occlusion, where only limited information such as a single body joint is available, the robot must continue tracking by leveraging whatever visual cues remain [56]. When these methods are insufficient and tracking is lost, re-identification becomes crucial to recover and resume following the correct person.

1) RE-IDENTIFICATION METHODS

Re-identification typically involves using different visual features to relocate the person of interest. Table 2 gathers the main method used in literature for re-identification in person following and guiding tasks. One common approach is the use of color histograms, where the robot identifies individuals based on specific color patterns, often linked to the clothing they are wearing [76], [87], [99]. Although straightforward, this method can be prone to errors in environments with similarly dressed people or varying lighting conditions. More advanced techniques include gait recognition, which analyzes the unique walking patterns of individuals to re-identify them [108], [115]. Gait recognition, however, requires reliable tracking of leg movements, which can be challenging in crowded or cluttered environments. Despite its complexity, this method offers robustness in cases where color-based identification is unreliable, providing an additional layer of discrimination. Machine learning approaches, especially those leveraging deep learning, have become increasingly prominent in person re-identification. Convolutional Neural Network have been particularly effective in learning robust visual features from a dataset of pre-collected images of the person, enabling the robot to distinguish them from other individuals in real-time [50], [64], [142]. These CNN-based solutions offer greater flexibility and accuracy by learning complex patterns and features less sensitive to lighting changes, pose variations, or occlusion. In addition to CNNs, online boosting [65], [101], [115] and attention-based [143] techniques have been employed for person re-identification, adapting to changing conditions and environments as the robot tracks the person over time. These methods continuously update the model based on new visual data, improving performance in dynamic settings where the person's appearance may vary.

2) HYBRID METHODS

A range of approaches here referred to as hybrid methods and summarized in Table 3, simultaneously address all three challenges: perception, tracking, and re-identification. Some of these approaches enable a wireless tethering solution with the robotic platform by utilizing wearable or transportable devices assigned to the user. These solutions provide continuous and reliable tracking even in complex environments where traditional vision-based systems might struggle. In one such example, the accompanying person wears an

TABLE 3. Hybrid perception-tracking methods.

Person detection sensors	Person detection data	Body part recognized	Features	Method	References
Ultrasonic, piezoelectric, acceleration sensors	Sensors linking	Wearable device with sensors	-	Wireless tethering	[149]–[151]
UWB anchors	UWB signal	Wearable UWB tags	-	Anchors triangulation	[152]–[154]
Exteroceptive multicamera	Color image	Marker (attached on person)	Marker pattern	Multicamera marker recognition, regression model	[40]
RFID sensors	RFID signal	RFID tag (hold by user)	-	RFID signal reading	[149], [155]–[158]
AOA sensors	AOA signal	AOA tag (hold by user)	-	Signal direction reading	[77]
Tactile supports, handles	Direct physical contact	-	-	Continuos localization of contact	[82], [117], [118], [134], [159]
Tether	Tether link with user	-	-	Tether following	[160], [161]

ultrasound sensor beacon. At the same time, the wheelchair is equipped with multiple ultrasound receivers [149] or piezoelectric ultrasonic transducers [150]. Figure 9 illustrates this configuration, where the wearable device continuously emits ultrasonic signals, which are detected by the receivers mounted on the wheelchair. Through the time-of-flight of the sound wave and triangulation between the multiple receivers on the wheelchair, it is possible to determine the precise position of the person. A different tethering approach is employed in [151], where the accompanying person wears an acceleration sensor. When localization with LRF is unstable due to obstructions, the acceleration sensor registers the kinematic assumed by the person. This data is then used to obtain an esteem of the person's position, facilitating the reconnection with the laser-based system.

Absolute localization systems represent another category of hybrid tracking solutions that can provide precise positioning data in fixed frame coordinates. These systems typically involve attaching beacons or tags to the subjects being tracked. At the same time, a network of anchors (such as fixed cameras, sensors, or receivers) is strategically mounted within the operational environment. The system can continuously and accurately localize the subject within the space by leveraging triangulation or time-of-flight calculations between the anchors and beacons. One notable advantage of this approach is its ability to track multiple subjects simultaneously, making it highly suitable for applications requiring concurrent monitoring, such as group navigation and collaborative robotics. For instance, UWB is known for its high accuracy in indoor environments, making it ideal for applications in confined spaces or crowded areas where GPS would be unreliable [152], [153], [154]. Similarly, as illustrated in Figure 10, exteroceptive multi-camera systems [35], [40] enhance tracking robustness by integrating visual data from multiple viewpoints. This multi-perspective approach reduces the risks of occlusions, significantly improving continuity and accuracy in dynamic environments. Radio Frequency Identification (RFID) systems are also widely used in such applications. In this approach, the followed person carries a unique RFID tag that continuously broadcasts an identifier signal. The robot can then track the individual by reading this signal using RFID readers integrated into its platform [149], [155], [156], [157], [158]. RFID offers the advantage of long-range detection and works well in environments where visual tracking might be compromised. Another related technique is the use of Angle of arrival (AOA) systems, which calculate the direction from which a signal is received to determine the location of the tagged person [77]. However, one significant drawback of absolute localization systems is their dependence on structured environments to function effectively. These systems require the installation of receiving anchors or cameras strategically placed throughout the area to facilitate accurate localization and tracking. This setup process can be complex, time-consuming, and costly, making it feasible primarily in large, controlled, and structured environments such as hospitals, warehouses, or industrial facilities. Unfortunately, this requirement limits the applicability of absolute localization solutions in smaller, unstructured, or constantly changing environments. Installing and maintaining the necessary infrastructure is often impractical or even impossible. For example, in outdoor settings or in buildings that cannot accommodate the required installation, the effectiveness of these systems is significantly reduced. Additionally, the need for a pre-existing, structured setup makes it difficult to use such methods in emergency situations or temporary locations where quick deployment is required.

In some cases, hybrid methods involve physical interaction between the robot and the guided person. For instance, some

FIGURE 9. The ultrasonic and piezoelectric tethering system, as described in [150]. In (a), both tethering components are shown: the green array attached to the wheelchair and the orange wearable beacon. In (b) and (c), the wireless tethering with the caregiver companion is illustrated.

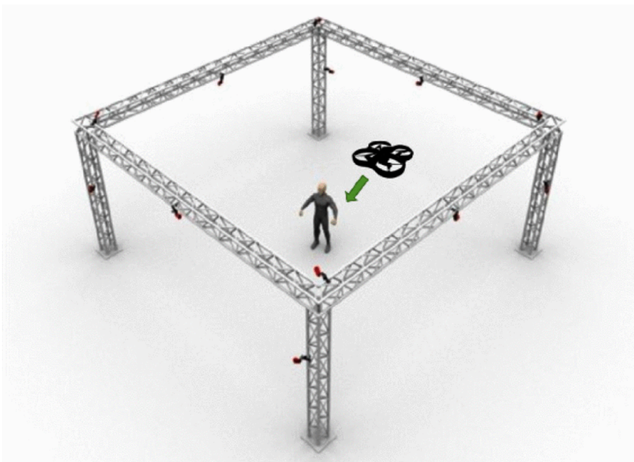


FIGURE 10. An illustration of an indoor multi-camera system, as described in [40]. Within the confines of the setup, both robots and humans equipped with specific markers can be tracked with exceptional precision.

healthcare and assistive robots are designed to maintain physical contact with the user, such as through handles [117], [134] or tactile supports [159]. These systems ensure continuous tracking and guidance, especially for users with mobility issues who require steady support while moving. Another related technique involves using tether interfaces, where a physical tether or leash connects the person to the robot [160], [162]. This approach ensures continuous localization of the user and provides a means to physically control the robot's movements. By maintaining a direct physical connection, the robot can track the user's position and adjust its behavior accordingly. In particular, the study conducted by [161] compares two different control methods for tethered robots: the Pseudo-joystick and the Follow the Leader control. The Pseudo-joystick method utilizes the tether's length and direction as steering input commands, allowing the user to control the robot's movements by manipulating the tether. In contrast, the Follow the Leader method treats the position of the tether tip as a dynamic goal that changes over time, with the robot continuously trying to reach this moving target. The study demonstrates a clear preference among users for the Follow the Leader method. This approach provides a more natural and intuitive experience, as users do not need to consciously manage the tether to steer the robot. Instead, they can move freely, knowing the robot will follow their lead. Additionally, this technique can be particularly beneficial in scenarios where users require both hands for other tasks, allowing them to rely on the robot for movement support without manual control.

C. PLANNING AND CONTROL

Historically, classical robotic navigation divides its architecture into two main modules: a planner and a controller, sometimes called local or motion planner. Global planners determine the optimal path from the robot's current position to a goal, given a map of the environment with known obstacles. Some commonly used global planner approaches include Breadth First Search (BFS), which systematically explores the environment, ensuring that the shortest path is found when the search is expanded uniformly. It is simple but can be computationally expensive in large maps due to its exhaustive nature [21]. A* search algorithm is a more refined method that uses heuristics to prioritize the search towards the goal, resulting in faster and more efficient path planning compared to BFS. A* is among the most popular planners in various navigation tasks due to its balance between optimality and efficiency [132]. Probabilistic Roadmap (PRM) randomly sample the environment and connect feasible configurations into a graph, which is then searched for a path. This method is effective for high-dimensional spaces or environments with many obstacles, where deterministic planners like A* might struggle [52]. Rapidly-exploring Random Trees (RRT) build a tree of feasible paths by incrementally expanding random samples from the robot's current position towards the goal. This approach is particularly advantageous in environments where paths are narrow or cluttered with obstacles [135].

Once the global path is generated, controllers take over to translate the path into velocity motion commands. These commands guide the robot while reacting to dynamic obstacles or environmental changes to maintain smooth and safe navigation. Some widely-used controllers include Proportional-Integral-Derivative (PID) controllers, straightforward feedback controllers that adjust the robot's velocity based on error values such as the distance from the path. They are often used for their simplicity and effectiveness in stable, linear navigation tasks [45], [68], [157], [163]. The Dynamic Window Approach (DWA) considers the robot's kinematic constraints, such as velocity and acceleration, to generate feasible trajectories in real-time. It evaluates multiple velocity options and selects the one that best balances obstacle avoidance, goal direction, and smooth motion [154], [164]. Model Predictive Control (MPC) algorithm uses a model of the robot's dynamics to predict future states and select optimal control actions. It accounts for both the current and anticipated future states, making it ideal for complex and dynamic environments where reactive control is needed [123], [136], [165], [166], [167]. Fuzzy logic controllers handle uncertainties and imprecision in sensor data by defining fuzzy rules rather than strict thresholds. This method allows for smoother and more adaptive control, particularly in complex or uncertain environments, which is why they are often used in HRI and tour-guide applications [18], [158], [168], [169].

In human following and guiding tasks, this two-module approach is often preserved. However, unlike fixed goal-based navigation, the goal here is dynamic, continuously shifting as the tracked individual moves through the environment. This dynamic goal introduces several complexities and unique challenges that require more sophisticated strategies to ensure smooth, effective, and socially compliant navigation. Figure 11 categorizes the primary methods discussed in this section for global planning and control, dividing them between traditional and social-oriented solutions.

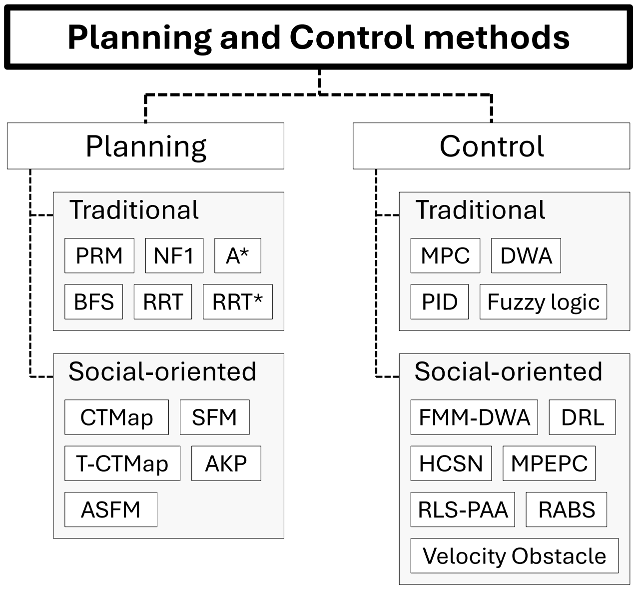


FIGURE 11. Categorization of planning and control methods. Socially-oriented methods are particularly well-suited for tasks involving human interaction, such as person following and guiding, where understanding and adapting to social cues is essential.

1) THE DYNAMIC GOAL PROBLEM

The primary challenge is adapting to the constantly changing goal as the person moves. The most straightforward approach is to precisely retrace the path taken by the followed person. In this method, the robot records the sequence of the person's positions over time and then follows this trajectory using either a simple pure pursuit algorithm [78] or a more sophisticated Robot Side method [74]. While this path-following strategy can be effective in static environments, it often fails in dynamic settings where previous clear paths may become obstructed. This limitation derives from the rigidity of the approach, which lacks adaptability to changes in the environment, such as new obstacles or crowded areas. An alternative strategy involves directing the robot toward the most recent known position of the tracked person. Although more straightforward in concept, this method provides the navigation system with greater flexibility. By focusing on the latest detected position rather than a rigid path, the robot gains the ability to dynamically adjust its movements, making it better suited for complex, unpredictable environments. As demonstrated in [15], this approach enhances the robot's adaptability and produces more human-like and socially compliant behavior. The robot can fluidly react to sudden changes in the environment, maintain smoother trajectories, and avoid awkward stops or rerouting that might disrupt the interaction with the person being followed. Furthermore, the second strategy inherently supports more socially acceptable navigation. By allowing the robot to operate with greater autonomy in choosing how to maintain proximity to the person, it can make context-aware decisions, such as avoiding close encounters with other pedestrians, preserving personal space, and smoothly maneuvering around obstacles. These factors contribute to a navigation experience that feels less mechanical and more aligned with how humans naturally move within shared spaces.

2) A HUMAN-AWARE NAVIGATION

The second essential aspect that the navigation system must consider during a following task is the social compliance of its movements. As discussed in Chapter II-D, socially acceptable navigation involves adhering to proxemics principles, which define the sensitive zones within a person's personal space. These principles ensures that the robot's proximity to humans remains comfortable and non-intrusive. Social navigation often involves dynamic adjustment based on real-time environmental factors and human behaviors. In this context, traditional planners and controllers must

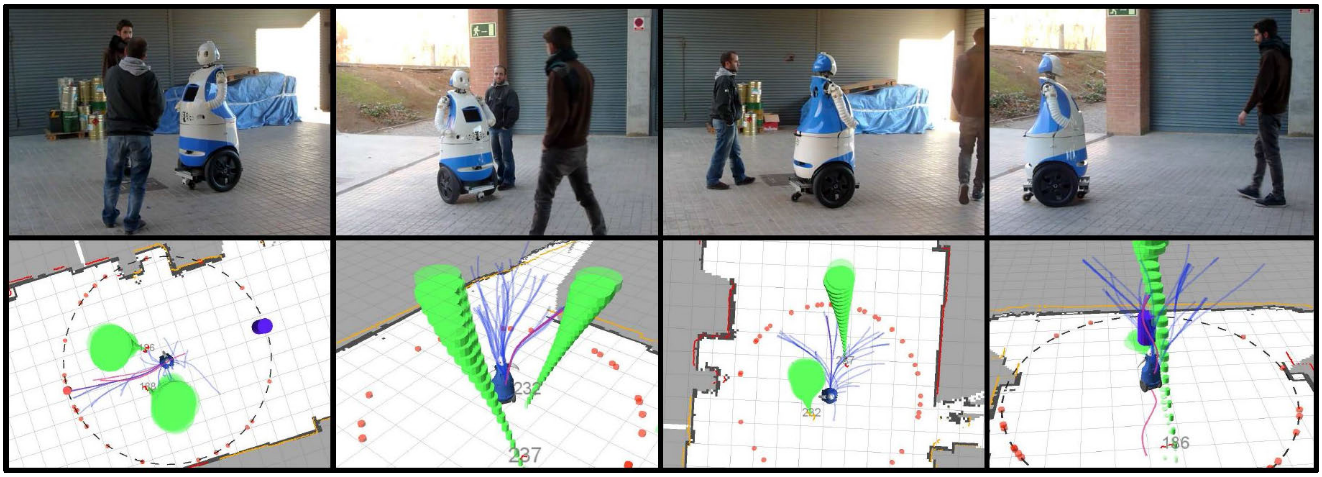


FIGURE 12. The Anticipative Kinodynamic Planner presented in [170]. The figures in the bottom row depict the *space* × *time* interface corresponding to the real-world scenario appearing above. In this interface, people are plotted as green cylinders and their predictions are drawn in the z axis, which correspond to time. The tree of paths calculated by the robot appears in blue and the best path is a red line.

be adapted to the appropriate social task. For example, in [171] and [172] the classical RRT* planner is integrated with Temporal Conditional Transition Maps (T-CTMap). The CTMap method, initially introduced in [173], is a grid-based representation that associates a probability distribution for an object exiting the cell, given its entry direction. In addition to the set of conditional probabilities, the T-CTMap stores a bivariate normal distribution to model the dependencies between entry and exit times. This method allows the robot to learn the average motion directions and speeds, as well as the speed variations. T-CTMap is particularly valuable in scenarios where robots must navigate crowded spaces like shopping malls or busy hallways. An improved version of the Dynamic Window Approach algorithm, the Fast Marching Method Dynamic Window Approach (FMM-DWA), is proposed in [77] to guide the robot towards the target individual. This approach considers the distance between the robot and its target and the alignment of the robot's heading with the direction provided by a direction gradient field, avoiding falling into local optima. Another hybrid solution integrates NF1-based global planning with an elastic band for path refinement and DWA for local control [174]. This setup mitigates typical NF1 issues like obstacle grazing and unsmooth trajectories, leading to more natural and fluid motion while following a person. MPCs are known for being particularly versatile and effective. The Model Predictive Equilibrium Point Control (MPEPC) framework [175] is an advanced online local trajectory planning and control algorithm. It allows the robot to dynamically adjust its trajectory to maintain a preferred distance and orientation relative to the person, a behavior often called ''person pacing''. Even classical PID controllers have been adapted for socially compliant following tasks. For example, PID controllers have been fine-tuned to maintain orientation toward the followed person [23] or to enable side-by-side navigation [30]. These adaptations involve adjusting control gains to prioritize smooth, natural interactions over strict path adherence, making the robot's behavior more human-like and comfortable.

In addition to traditional methods, research has increasingly focused on developing novel and socially acceptable navigation strategies specifically tailored for human following and guiding tasks. A prominent example is Human-Centered Sensitive Navigation (HCSN), first introduced in [176], which prioritizes human-centric factors by introducing different rules to regulate robot behaviour. HCSN leverages a combination of sensor data, predictive modeling, and human behavior analysis to optimize both the safety and smoothness of the robot's movements. Another advanced approach is the Robust-Adaptive-Behavior Strategy (RABS) presented in [18]. RABS is based on a fuzzy inference mechanism, which dynamically adjusts the robot's velocity based on three key inputs: the robot's current speed, the Highest-Traversability-Score Direction (HTSD), and the Adaptive-Safe-Following Distance (ASFD). This strategy allows the robot to maintain a safe and comfortable distance from the followed person while adapting its behavior to changing environmental conditions and user preferences. The Recursive Least Square Parameter Adaptation Algorithm (RLS-PAA), explored in [153], achieves smooth and socially compliant side-by-side navigation. This approach tunes the robot's behavior based on real-time sensor data, continuously adapting navigation parameters to balance safety, efficiency, and user comfort.

The Anticipative Kinodynamic Planner (AKP) is a powerful planner largely employed in following and guiding applications. Presented for the first time in [170], the AKP is a motion planning strategy primarily designed for human-robot cooperative navigation, which focuses on predicting the future states of both the robot and surrounding dynamic agents (e.g., humans) and integrates this foresight into the planning process. This method allows the robot to anticipate the people's movements and other obstacles, enabling smoother, safer, and more socially compliant navigation [177], [178]. Unlike simple controllers that only consider the robot's kinematics, the AKP also considers dynamic constraints, such as acceleration, velocity limits, and control inputs. As can be seen in Figure 12, the AKP predicts the future trajectories of humans in the environment by using models based on human behavior patterns. By considering the predicted trajectories of moving agents, the AKP selects paths that minimize potential conflicts, leading to smoother interactions and better adherence to social norms. In [12], the model used by AKP for predicting future human trajectories is the Social Force Model (SFM). Maybe one of the most famous models for social-aware navigation, the Social Force Model [179] is a widely-used concept in robotics and crowd dynamics that models the interactions between individuals and their environment using a physics-based approach. Developed initially by Dirk Helbing and Péter Molnár in the 1990s, the model describes how pedestrians navigate their surroundings by considering social and physical forces. These forces influence the acceleration and direction of a person's movement and are derived from both the environment and other individuals. The model distinguishes between repulsive and attractive forces. Repulsive forces are those between individuals and with obstacles, ensuring to avoid collisions. On the other hand, the robot is driven by attractive forces towards the goal. In situations where individuals move in groups, attractive forces also maintain group cohesion and keep them together. The Social Force Model has been extensively exploited for following and guiding applications to refine the planner's path [132] for different case studies. For example, in [133] it is used to guide a specific person in a crowded environment, avoiding collisions with other people. Another important advantage the Social Force Model gives is its ability to be integrated and adapted to different custom situations. For example, it can be tailored to follow a dynamic goal [127], to maintain the angle and distance between the robot and the user and obtain side-by-side navigation [131], [135], [180], to avoid occluded areas [181], and to adapt to flying robotic platforms like drones, with the Aerial Social Force Model (ASFM) variant [40].

Even with the most advanced tracking algorithms discussed in Chapter II-B, challenges arise in reliably following a target person if the navigation system does not fully integrate with perception systems. Therefore, significant research has been dedicated to developing control architectures that offload some responsibilities from perception systems to enhance overall performance. A common approach in social navigation control is to decouple linear and angular velocities, optimizing the robot's orientation to keep the target within the camera frame as much as possible [9], [23], [53], [67], [74]. For example, in [147], an omnidirectional robotic platform employs a traditional DWA controller to manage plane linear velocities for following the target, while angular velocities are handled by a Deep Reinforcement Learning (DRL) agent. As can be seen in Figure 13,

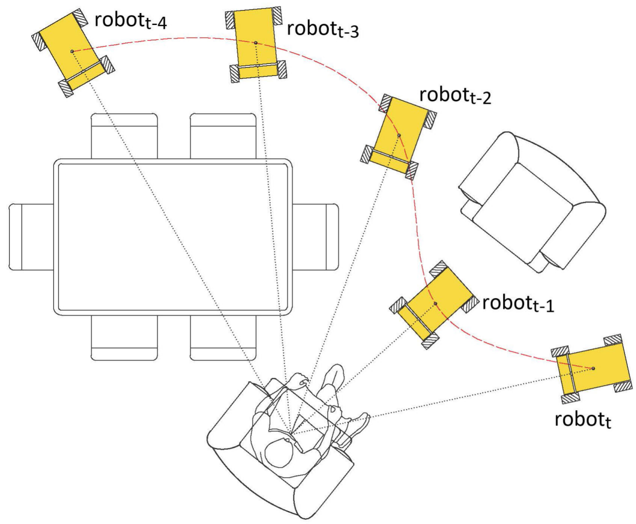


FIGURE 13. By decoupling linear and angular velocity, omnidirectional robotic platforms can follow a user while consistently keeping them within the camera's field of view [9].

this hybrid approach allows the robot to maintain its orientation toward the user even while navigating around obstacles, which is particularly advantageous compared to differential-drive robots that, instead, require changes in orientation to follow curved paths. Another approach involves using a motorized pan-tilt camera to keep the person in view without rotating the entire robot [81], [109], or using an omnidirectional camera [6], [20], [36]. This strategy enables the robot to continuously track the target while maintaining stable navigation, enhancing the system's robustness and responsiveness in dynamic environments.

Similarly to the perception and tracking cases, Machine learning methods have increasingly been adopted as effective and versatile controllers for human following and guiding tasks. While supervised solutions have been proposed [168], Deep Reinforcement Learning methods are usually the most adaptable to end-to-end navigation tasks [182], [183], [184]. For instance, a DRL-based framework in [185] has been developed specifically for both human following and guiding scenarios. Three different agents are employed to progressively enhance the robot's navigation abilities, where the most basic one receives the goal position and 2D laser scan data as input. A second agent is provided with additional safety zone information for each obstacle to make the navigation safer, whilst a third and most advanced agent incorporates additional detailed semantic social states as input, such as type, social status, radius, safety distance, distance to the agent, and position in the agent frame of each detected person. This enriched input allows the agent to better understand social dynamics and adjust its behavior accordingly. A dense reward function returns positive rewards when the agents move toward or reach the target with a reasonable number of steps. Negative rewards are only given for collisions or if the agents get too close to a static or dynamic obstacle. The author demonstrated that providing additional social semantic information to the agent leads to better following and guiding navigation results.

D. INTERACTION

Human-Robot Interaction (HRI) has historically received less attention from roboticists, who primarily focused on technical challenges related to perception, navigation, and control. However, as robots started being integrated into environments shared with humans, particularly with the rise of social robotics, the significance of effective interaction modalities, namely how robots approach, communicate, and behave with humans, has become widely recognized. This shift is particularly critical in applications such as person following and guiding, where robots have to work closely with people.

Interaction can then be divided into two key aspects: social behavior, which refers to how the robot navigates and behaves while interacting with humans, and human-robot communication, which encompasses the methods the robot uses to receive input from users and convey information to them. Both elements are crucial for ensuring effective and natural human-robot interactions.

1) SOCIAL BEHAVIOUR

Recent research has emphasized the importance of designing robots capable of understanding and adapting to human social cues, preferences, and behaviors. The ability of robots to maintain socially appropriate distances, adapt their speed, and engage in non-verbal communication are now recognized as key components in delivering a positive user experience during HRI [186], [187]. For example, person following robots must not only track and follow a human accurately but also maintain a distance that feels comfortable and non-intrusive. This requires sophisticated models of proxemics—how humans perceive and manage space in social contexts. These models are informed by both psychological principles and real-world experimentation, enabling robots to adjust their behavior based on user feedback, environmental conditions, and task requirements. One of the emerging areas of focus in HRI is the dynamics of role assignment, particularly in leader-follower relationships during collaborative tasks. In these scenarios, robots and humans must constantly negotiate roles, determining who takes the lead and who follows. Studies have explored how robots can adapt their behavior based on real-time cues from humans, such as changes in pace, direction, or posture. Adaptive behavior models allow robots to transition between leader and follower roles based on the context, which improves both task efficiency and user satisfaction [188]. For instance, if a robot detects that a human is slowing down or stopping, it can interpret this as a hint to take the lead temporarily or adjust its following behavior accordingly.

Side-by-side navigation represents a key step toward enhancing HRI by making it more engaging, intuitive, and socially aware. Traditionally, most approaches have treated side-by-side navigation as a following task, where the robot's primary objective is to track and follow the user, dynamically adjusting to their speed and direction [135], [150]. In this traditional model, the robot assumes a subordinate role, adapting its behavior to follow the human leader, which limits the interaction to a reactive and hierarchical framework. Recent research has increasingly focused on evolving this interaction into a more collaborative model, where the roles of follower and leader are fluid and interchangeable. This shift allows the robot and the human to walk side by side as peers, switching roles dynamically based on context or environmental conditions [130]. In this model, the robot is not just a passive follower but an active participant in the navigation process. In fact, as can be seen in Figure 14, the robot is capable of assuming leadership when necessary, such as when navigating through narrow pathways or crowded environments. This adaptive role-switching enhances the fluidity of interaction and results in smoother, more intuitive navigation experiences, closely resembling how two humans would naturally walk together. This aspect is particularly critical in sensitive applications such as healthcare, where robots are increasingly used to assist patients. In this context, robotic platforms like assistive wheelchairs benefit immensely from side-by-side navigation. By incorporating autonomous following alongside peer-based navigation, these systems reduce the caregivers' physical and cognitive burden while promoting more meaningful interactions between the patient and caregivers [25].

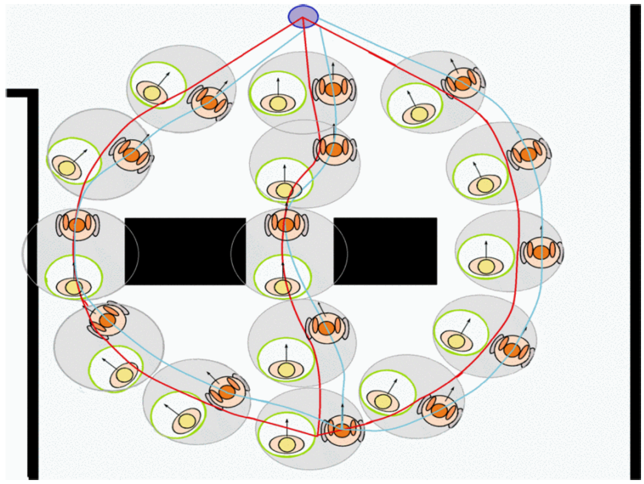


FIGURE 14. Side-by-side navigation demands a deeper understanding of social dynamics. In this illustration, three navigation solutions for human-robot side-by-side navigation are presented. While the first path offers a valid solution and the second represents the fastest and shortest route, only the third path successfully maintains the side-by-side formation, ensuring smoother and more socially appropriate interaction [135].

The potential of side-by-side navigation extends beyond healthcare into various domains where close human-robot collaboration is essential. For instance, in environments like shopping malls or parks, companion robots can accompany users in a natural and human-like way, enriching the overall user experience [12], [126]. Similarly, in collaborative manufacturing, the ability to switch roles based on task

(a) Pointing to a whole exhibit (b) Pointing to part of an exhibit

FIGURE 15. Through arm movements, robots can communicate and guide users [155].

requirements—such as leading when carrying tools or following when monitoring—can lead to more seamless and productive collaboration [189], [190]. Another critical component of side-by-side navigation is the development of robust human intention prediction models. Understanding and predicting a human's intended path or movements allows the robot to adjust its trajectory, proactively enhancing safety and user comfort. Research has shown that incorporating human intention modeling, such as through motion prediction algorithms and behavioral cloning, can significantly improve the fluidity of interaction in dynamic environments where sudden changes in pace or direction are common [191].

2) HUMAN-ROBOT COMMUNICATION

Communication between humans and robots can occur through various modalities. Among these, vocal communication stands out as one of the most intuitive and familiar methods, given that speech is humans' primary mode of communication. In robotics, speech synthesis is typically achieved through text-to-speech (TTS) technologies, enabling robots to interact verbally with users. This capability allows robots to confirm received commands, provide feedback, and offer verbal guidance regarding tasks or directions [67], [156]. For instance, robots functioning as tour guides often rely on speech to deliver explanations about museum exhibits or to keep users informed about the progression of the guided tour [155], [158]. In addition to speech synthesis, many robotic platforms are equipped with sophisticated speech recognition systems that enable robots to engage in meaningful conversations with humans [117], [118]. In recent designs, these systems are powered by advanced natural language processing (NLP) algorithms, allowing robots not only to understand spoken language but also to respond appropriately. Beyond simple conversational exchanges, speech recognition is an efficient command input method, allowing users to control robots through voice commands. Vocal communication is particularly useful in hands-free or accessibility-focused scenarios, where traditional input methods like touchscreens or remote controls may be impractical [10], [81]. Speech-based control is increasingly integrated into service robots, personal assistants, and even assistive technologies for individuals with disabilities, enhancing usability and inclusivity.

Another widely adopted mode of communication between humans and robots is through graphical interfaces displayed on screens. These interfaces provide a visual medium for conveying information and facilitating interactions. In some cases, a simple screen is used to communicate the robot's intentions, status, or upcoming actions through visual or text-based messages, enhancing transparency in robot behavior [82]. However, more advanced systems commonly incorporate touchscreens. Touchscreens allow users not only to receive information but also to input commands, adjust settings, and interact with the robot in a more intuitive and direct manner [36], [71].

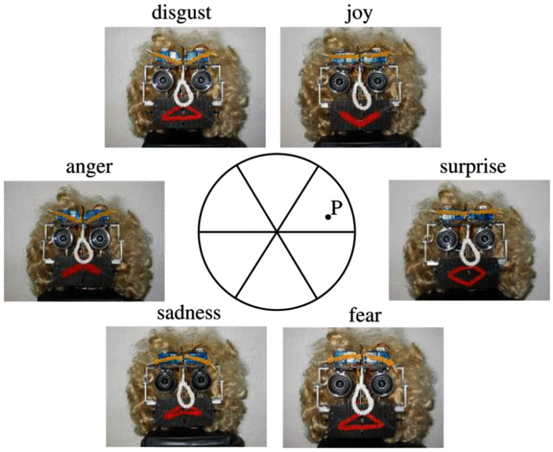


FIGURE 16. Many social robots are designed with expressive faces to enhance communication through emotional cues. For example, Alpha [69] is equipped with motorized eyes, eyebrows, and mouth, enabling it to display a range of facial expressions.

Gesture recognition is another important communication method that allows robots to understand and interpret commands through human body language. For example, robots can detect a pointing gesture to determine a desired goal [109] or recognize a hand or arm signal to execute a specific command [34], [58], [67], [124]. This capability is especially useful in scenarios where verbal communication may not be feasible or preferred [192]. Beyond receiving commands, robots can also use gestures to provide feedback to users. As illustrated in Figure 15, this communication method is often employed in environments where interpersonal interaction is key, such as museums, healthcare settings, or customer service roles [156], [193]. Furthermore, gesture recognition is often supplemented by multimodal communication strategies, such as combining speech with graphical interfaces or facial expressions to enhance the clarity and effectiveness of communication (Figure 16). This multimodal approach enriches the user experience and allows the robot to adapt its communication style based on the context or user preferences [69], [158].

Table 4 summarizes the primary human-robot communication modalities discussed in this section, highlighting the key application areas where each is commonly employed.

III. APPLICATION FIELDS

The robot's intended application plays a crucial role in determining its design features, sensor suite, interaction models, and overall user experience. In healthcare and

TABLE 4. Communication methods.

Communication modality	Input modality	Output modality	Task	References
Vocal, gestures	Hands gestures	text-to-speech	Person follower	[67]
Vocal, gestures	Given remotely by human	text-to-speech, arms gestures	Mall guide	[156]
Vocal, gestures	-	text-to-speech, arms gestures	Tour guide	[155]
Vocal	Speech recognition	text-to-speech	Healthcare assistant	[117], [118]
Vocal	Speech recognition	text-to-speech	Person follower	[81]
Vocal, remote device	speech-to-text, remote device	text-to-speech	Personal assistant	[10]
Graphical interface	Haptic interface	LCD screen	Robotic walker	[82]
Graphical interface, vocal	Touchscreen	Touchscreen, text-to-speech	Shopping guide	[36]
Graphical interface, vocal	Touchscreen	Touchscreen, text-to-speech	Tour guide	[71]
Gestures	Hands, arms gestures	-	Person follower	[34], [58], [109], [124], [192]
Gestures	Hands gestures	-	Tour guide	[193]
Vocal, gestures	Speech recognition	text-to-speech, face emotional response, arms gestures	Tour guide	[69]
Vocal, gestures, textual	Textual	text-to-speech, face emotional response	Tour guide	[158]

personal assistance, for example, especially when serving vulnerable populations such as the elderly or disabled, the emphasis is on physical safety, intuitive interfaces, and adaptive behavior. These robots must be user-friendly and capable of understanding human intent, emotional states, and environmental context, while maintaining autonomy and strict safety standards. Social acceptance is equally important, as service robots interact closely with humans and must demonstrate socially appropriate behavior.

Conversely, robots in industrial settings, warehouses, and logistics prioritize durability, operational efficiency, and load-bearing capacity. These platforms tend to be more utilitarian, focusing on functionality and productivity rather than aesthetics or human interaction. Since the primary goal is efficient task execution, there is little need for social engagement, and the design focuses more on robustness and performance than on fostering human-centered interactions.

The application also influences the robot's role in supporting users. For instance, robots tasked with guiding individuals unfamiliar with their surroundings or those with mobility impairments often serve as escorts, leading users to their destinations. Examples include guide robots in airports or museums, where the robot navigates and provides information and instructions. In cases where users are already familiar with the environment but require physical assistance, such as carrying heavy loads or tools, the robot's role shifts from guiding to following and supporting. Such configurations are commonly seen in warehouse robots that follow workers, carrying equipment or goods, and in home robots that assist with daily tasks.

The following sections will categorize literature based on these application fields, identifying key innovations and highlighting how they address the unique challenges associated with each domain.

A. HEALTHCARE

The field of service robotics in healthcare is extensive and diverse, encompassing a wide range of applications, from supporting medical staff in hospital settings to providing personal assistance to the elderly and physically impaired in their daily lives. Figure 17 categorizes and highlights the primary applications within this domain, illustrating the diverse roles that robotics play in enhancing both healthcare delivery and quality of life for patients.

1) MEDICAL ASSISTANCE

In hospitals, service robots have become indispensable in modern healthcare, significantly enhancing operational efficiency by taking over routine utility tasks traditionally

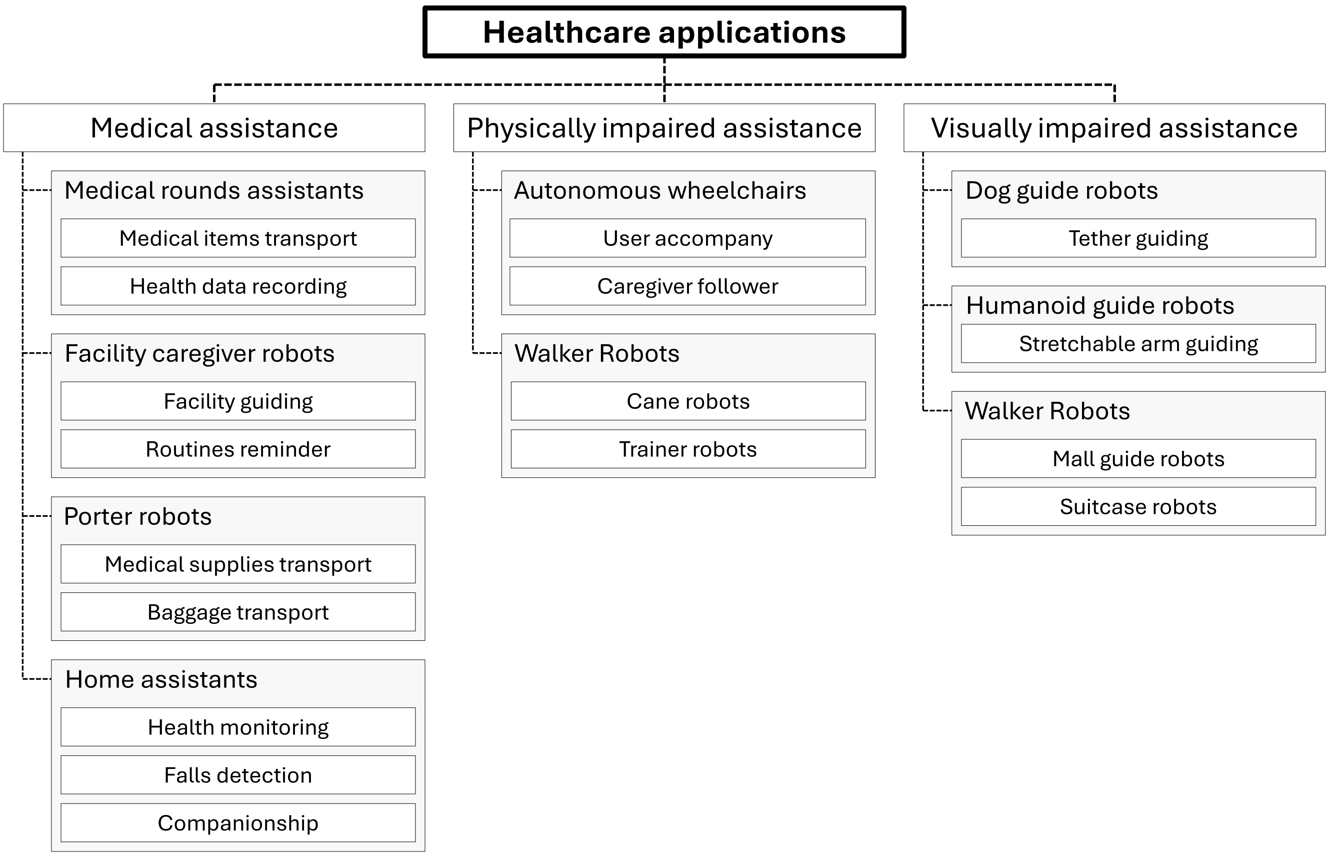


FIGURE 17. Classification of person following and guidance applications in the healthcare sector.

managed by nurses and healthcare staff. As illustrated in Figure 18, these robots are designed to handle a wide range of responsibilities, such as transporting medical supplies, patient files, and medications, which alleviates the burden on medical professionals and allows them to focus more on direct patient care and critical decision-making. Fully autonomous robots, for instance, can accompany nurses during medical rounds, carrying essential items and assisting with tasks like recording electronic health data in real-time [3], [151]. This not only streamlines workflow but also reduces the likelihood of human error, contributing to improved patient outcomes and overall quality of care. Several robotic platforms have been specifically developed to tackle pandemic-related challenges, particularly during the COVID-19 outbreak. Such robots autonomously disinfect high-contact surfaces in public spaces by emitting UV-C light, which effectively neutralizes germs, bacteria, and viruses by breaking down their DNA and RNA [194]. Additionally, many of these robots are equipped with thermal cameras to remotely measure the body temperature of individuals entering medical facilities, allowing for early detection of potential symptoms while minimizing human contact.

In assisted living facilities and other structured healthcare environments, robotic platforms play a vital role in supporting individuals with cognitive or physical limitations, enhancing both their quality of life and the efficiency of care provided by staff. These robots can assist residents by guiding them to various locations within the facility, such as dining areas, recreational spaces, or therapy rooms [117]. Beyond navigation, they remind residents to perform essential daily routines, such as taking medications, staying hydrated, and eating meals at regular intervals. These reminders can be tailored to each individual's schedule and medical needs, ensuring that crucial tasks are not overlooked [118]. Such integrations are particularly valuable in managing chronic conditions, offering continuous monitoring that reduces the risk of emergency situations and promotes proactive care.

Autonomous robotic personal assistants can assist patients by transporting various medical equipment, such as medications, drip stands, and other necessary supplies, reducing the workload of healthcare staff and improving overall efficiency [4], [39]. One particularly valuable application of robotic personal assistants is supporting patients undergoing Home Oxygen Therapy (HOT). HOT is a critical treatment for individuals with severe lung diseases who require continuous oxygen administration via a tank. Managing and transporting these oxygen tanks can be challenging, especially for patients with limited mobility. Robotic platforms have been developed to carry oxygen tanks and follow patients as they move

(a) (b) (c) (d)

FIGURE 18. **Medical assistance robots take on various forms and designs tailored to their specific tasks. For example, Terapio [3], shown in (a), is designed as a medical rounds assistant, while Lio [194] in (b) serves as a caregiver in assisted living facilities. The robots in (c) and (d) are examples of porter robots: the first [161] supports Home Oxygen Therapy, with the tether following system, and the second [123] functions as a home assistant capable of carrying baggage and luggage.**

through both indoor and outdoor environments. These robots often utilize tether interfaces to detect and track the relative position of the patient, ensuring the oxygen supply is always within reach [160], [161]. This method allows patients to maintain mobility and independence while receiving essential medical care. In addition to medical support, robotic personal assistants can significantly improve the quality of life for people with mobility difficulties by assisting with everyday tasks that might otherwise be difficult. For example, carrying baggage and luggage can be a strenuous activity for these individuals. Robotic assistants with carrying capabilities can take over this burden, making travel and daily activities more accessible and less physically demanding [41], [123].

Robotic personal assistants are also increasingly being adopted in home environments, especially for elderly and physically impaired individuals who require continuous support. These robots offer vital assistance that can significantly extend the period individuals can remain in their homes, delaying or even preventing the need for transitions to retirement homes or healthcare facilities. By providing help with daily tasks like moving around the house, fetching items, and reminding users of scheduled activities, robotic personal assistants enhance the users' autonomy and improve their overall quality of life. One of the critical functions of home robots is monitoring for dangerous situations, such as slips, falls, or sudden health emergencies [195]. Upon detecting such incidents through built-in sensors or AI algorithms, the robots can promptly alert caregivers or emergency services, ensuring swift intervention. More advanced models incorporate predictive analytics, capable of identifying high-risk behaviors or environmental factors, such as frequent stumbling or cluttered spaces, which could lead to accidents [73]. This solution allows the system to issue preemptive warnings or even make adjustments, like suggesting safer pathways or recommending assistive measures, thereby actively preventing incidents before they occur. In addition to physical assistance, many home-based robots come equipped with natural language processing and voice interaction capabilities, enabling them to engage users in conversation, offer medication reminders, and even provide emotional support. By serving as companions, these robots address the social and psychological needs of their users, reducing feelings of loneliness and isolation, which are common among elderly individuals [10].

2) PHYSICALLY IMPAIRED ASSISTANCE

In terms of physical assistance, robots serve as crucial aids for guiding and supporting individuals with mobility challenges, a role that has traditionally been filled by caregivers. For instance, robots equipped with actuators and stabilization mechanisms can help patients stand, walk, or transition between different positions, such as moving from a bed to a wheelchair. These tasks, which are often physically demanding for caregivers, can be executed more consistently and safely by robotic systems, thereby reducing the risk of injury for both caregivers and patients. One class of such devices, often called walker robots, is designed specifically to assist users with walking. As can be seen in Figure 19(a), these robots typically feature sturdy handlebars for users to grip and support themselves while the autonomous platform navigates to guide them towards the designated location [117], [118], [134], [196]. Similar to other advanced healthcare platforms, walker robots can offer continuous health monitoring during use. Tracking vital signs and other health metrics can provide real-time feedback and alerts, ensuring that the user remains in good health and receives prompt assistance if needed [79]. Some walker robots are also equipped with display screens showing to the user the intended direction of movement, helping them navigate to their destination with greater confidence [82]. These visual aids make the system more user-friendly, especially for individuals with cognitive impairments or those unfamiliar with robotic assistance.

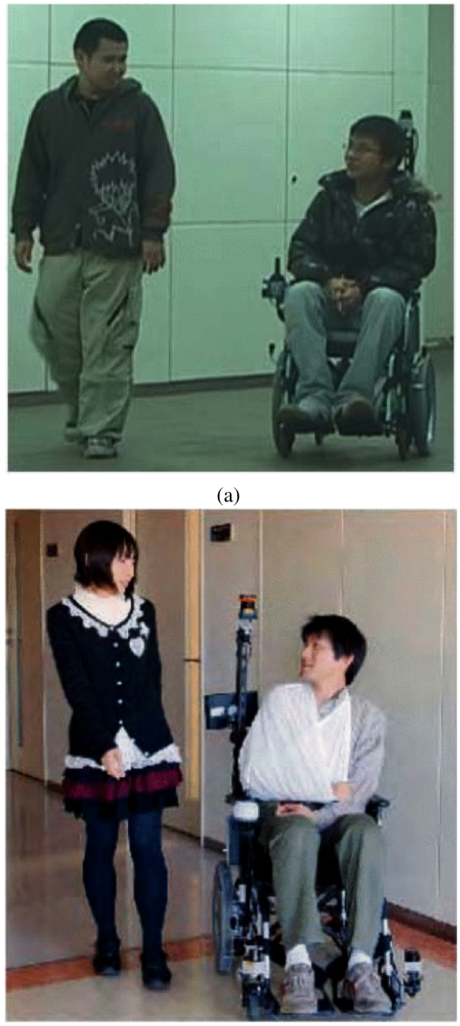
Another alternative to traditional walkers is the cane-robot, a smart and autonomous platform designed to function as an enhanced walking cane. Figure 19(b) shows a prominent example of this kind of platform. Unlike conventional passive devices that rely entirely on the user's input to determine speed and direction, cane-robots leverage sophisticated technologies to offer a more interactive and supportive

(a) (b) (c)

FIGURE 19. Various types of robotic walkers are here illustrated. In (a), a standard walker robot is shown designed to fully support a physically impaired user [79]. In (b), a typical smart cane robot [8]. Finally, in (c), a proprioceptive reference point robot is depicted, capable of stabilizing gait for elderly individuals and stroke patients [159].

experience. By detecting the forces and torques applied to the cane's handle, these robots can adapt their behavior in real-time to meet the user's needs. One of the key features of cane-robots is their ability to utilize follow-ahead functionality, enabling the robot to anticipate the user's movements with human intention estimation methods [8], [45], [163]. Cane-robots are equipped with sensors to navigate complex environments, detecting and avoiding obstacles the user might encounter. Applying gentle yet noticeable steering forces through the handle provides intuitive force feedback, helping users maneuver through unfamiliar or cluttered spaces with greater ease and confidence. This method not only enhances safety but also boosts the user's autonomy and independence. In addition to obstacle avoidance, cane-robots can also guide users towards specific destinations. They achieve this by balancing two critical factors: adapting to the user's preferred speed and applying directional forces through the handle to lead them toward the desired location. This dual capability ensures safe and efficient movements, making cane-robots versatile tools for various settings and mobility needs. Beyond these active systems, researchers have explored simpler solutions aimed at providing just enough support to stabilize gait without requiring significant force or coordination from the user. Instead of relying on a handle, robots with a fingertip contact plate offer a light-touch reference point that elderly individuals or stroke patients can use to enhance their balance while walking (Figure 19(c)). This proprioceptive solution helps maintain stability during overground gait without the need for full weight support or extensive user input [159]. Such designs focus on minimal yet effective intervention, making them especially beneficial for users who require a subtle form of guidance rather than full physical assistance.

In rehabilitation settings, walker robots are increasingly used not only as mobility aids but also as active participants in therapy. They engage patients in walking exercises and self-guided training routines, crucial for improving mobility, balance, and spatial orientation skills [87], [197]. By offering timely corrections and personalized coaching, they help patients maintain proper form, which is crucial for effective rehabilitation and faster recovery. Additionally, these robotic systems can track progress over time, adapting exercise routines based on the patient's improvement, thus offering a tailored rehabilitation experience that optimizes outcomes. Beyond physical assistance, the interactive nature of these robots stimulates patient engagement and motivation, making the often challenging rehabilitation process more accessible and encouraging long-term adherence to recovery plans.



(a)

(b)

FIGURE 20. Side-by-side following for autonomous wheelchairs is an effective solution that enables caregivers to interact more easily with the patient. Examples of such systems in indoor environments are shown in (a) and (b), as presented in [25] and [26], respectively.

Autonomous person following systems have been increasingly integrated into electric wheelchairs, to improve user mobility and provide enhanced caregiver support. Given that many wheelchair users are often accompanied by healthcare personnel, the design of these systems must prioritize reducing their workload while allowing communication between the followed person and the wheelchair user. To address these needs, recent research has shifted towards side-by-side wheelchair accompaniment rather than the traditional leader-follower configuration. In side-by-side systems, the focus is on creating a peer-to-peer interaction dynamic rather than a hierarchical one, where the caregiver acts as a leader and the wheelchair merely follows. The goal is to make navigation feel like a cooperative experience, reducing the robot's intrusiveness and enhancing social acceptability. As can be seen in Figure 20, this approach not only improves user comfort but also supports better engagement between the user, caregiver, and surrounding pedestrians. For instance, the caregiver is given the flexibility to actively propose and execute navigational plans, making the interaction more natural and intuitive, as highlighted by works like [130]. The autonomous wheelchair, in turn, adapts its behavior to match the speed and direction assumed by the followed individual while maintaining an appropriate distance and orientation. This ensures that both the wheelchair user and the caregiver are comfortable during the interaction, allowing for smooth conversations and unobstructed movement. For example, studies have shown that adapting to the caregiver's walking speed and maintaining a proper lateral distance is critical for user comfort and natural interactions [150], [175]. In dynamic and crowded environments, the control system of these wheelchairs needs to be responsive to sudden movements, stops, or changes in direction by the caregiver. The ability to quickly adjust to such situations is crucial for safe and effective navigation, particularly in unpredictable settings such as busy hallways or outdoor spaces. Predictive algorithms that anticipate harsh movements and sudden stops allow autonomous wheelchairs to navigate smoothly in environments with high levels of human activity [42], [144].

While many studies focus on indoor environments like hospitals and care facilities, there is also growing interest in adapting these autonomous systems for outdoor use and in complex public spaces such as museums, railway stations, and shopping centers. Navigating such cluttered and crowded areas poses additional challenges, such as dealing with varying terrains, moving obstacles, and fluctuating environmental conditions. Research in this area has demonstrated the importance of robust perception systems and adaptive control strategies that can handle these complexities while maintaining a comfortable and safe user experience [23], [198]. Interaction with the environment is another key consideration. For example, when the caregiver stops to engage with an object, person, or point of interest, the wheelchair must recognize this pause and autonomously move to a suitable position. This behavior enhances the social acceptability of the platform and reflects a deeper understanding of human social norms. In these scenarios, the wheelchair must adjust its orientation and distance to facilitate interaction while avoiding obstructions and maintaining visibility for both the patient and caregiver. This adaptability ensures that the robotic platform blends into everyday activities, improving user satisfaction and social acceptance [25], [26].

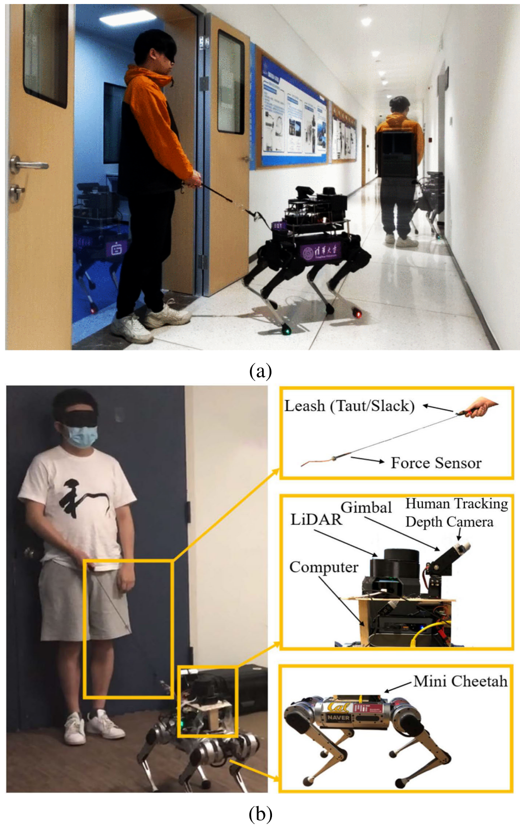


FIGURE 21. Robotic assistants for visually impaired users in indoor environments. In (a) [199] and (b) [200], two typical quadrupedal guide dog robots.

3) VISUALLY IMPAIRED ASSISTANCE

Another significant challenge in the healthcare field is providing assistance to visually impaired individuals. While guide dogs offer the highest degree of mobility and independence, their training is time-consuming and expensive, requiring selective breeding and specialized training programs. As a result, robotic assistants have emerged as viable alternatives, offering a scalable and potentially more accessible solution. This task demands a multifaceted approach, incorporating advanced technology, user-centered design, and robust safety measures to ensure the robot can provide reliable and effective support. One primary challenge is ensuring that robots can navigate complex environments safely and efficiently. Visually impaired individuals rely heavily on the robot's ability to detect and avoid obstacles, which may include both stationary objects, such as walls and furniture, and dynamic elements, such as moving vehicles, pedestrians, and pets. Additionally, these robotic assistants must be capable of operating seamlessly in both indoor and outdoor environments, each presenting unique challenges. Indoor navigation requires the robot to manage narrow hallways, sharp turns, and varying floor surfaces, while outdoor navigation involves dealing with uneven terrains, curbs, weather conditions, and interpreting traffic signals.

The earliest examples of robots designed for this purpose took inspiration from the natural world, emulating the behavior of guide dogs. Guide robots could detect and avoid obstacles while guiding the user using a tether or leash system [75], [122]. This approach provided a familiar and intuitive interface for users accustomed to guide dogs, making the transition to robotic assistance smoother. Recent advancements have extended this concept to include quadruped robots [199], [200], [201], as illustrated in Figure 21, as well as platforms that facilitate more human-like interactions, focusing on a closer, more personal form of guidance. For instance, some modern robotic assistants are designed to lead the visually impaired user by hand, using a stretchable arm interface while maintaining an appropriate distance. This method not only allows for more precise guidance but also enables the robot to perform additional functions, such as providing information or explanations, which can be particularly useful in settings like museums or exhibitions [66], [74]. This physical human-like accompaniment can enhance the user's experience by making the interaction feel more natural and supportive [202]. Walker robots equipped with handles have been extensively employed also to assist visually impaired users in various settings. These robots can accompany users in grocery stores and supermarkets, helping them navigate aisles, find products, and even carry purchased items [80]. Another innovative solution, involves autonomous suitcase-shaped robots capable of guiding users to their destination while avoiding obstacles [83]. These robots are particularly beneficial for travel and daily errands, offering independence and ease of movement. In all these cases, control over walking speed is shared, requiring the robot to adapt to the user's pace to ensure smooth and comfortable guidance. Walker robots often include a clutch mechanism to provide flexibility and allow users to manually push or pull the robot when necessary. This feature disengages the motorized wheels from the motor, enabling manual control and ensuring that users can easily maneuver the robot in various situations.

B. LOGISTICS

The application of robotic agents in warehouses and logistics facilities has seen rapid growth, driven by advancements in automation technologies, increasing demand for operational efficiency, and the need to address persistent labor shortages. Among these innovations, human following and leading robots are becoming integral to modern factories, where they reduce the physical strain on workers by taking over repetitive, labor-intensive tasks such as grasping, lifting, and transporting heavy tools, materials, and products. These robots not only alleviate the physical burden on workers but also enhance overall productivity, safety, and workflow management in these settings.

Human following porter robots are particularly valuable in dynamic environments where workers are constantly on the move. These robots can autonomously trail workers, carrying tools or parts, freeing up the human workforce to focus on more skilled tasks. For example, in assembly lines or large manufacturing floors, porter robots can seamlessly integrate into the flow of operations, ensuring that necessary items are always within reach while minimizing downtime [203]. Additionally, leading robots can guide workers through optimized paths in warehouses or storage areas, ensuring efficient task completion and reducing time wasted in transit. These platforms typically take the form of autonomous carts equipped with large tray-like surfaces where heavy components can be placed and moved around the facility [5]. Some advanced models go beyond basic load carrying by integrating robotic arms that allow them to pick up, transport, and place objects like a human would handle them. These collaborative robots, often referred to as cobots, are increasingly used in tasks like inventory management, order picking, and assembly line support [191]. In smart factories and logistics centers, robotic systems are versatile enough to handle a range of tasks. They are used to transport tools for maintenance, finished products, subsidiary materials, and other objects essential to production and warehouse operations. Moreover, robots can be deployed both indoors and outdoors, supporting tasks like assembling bulky objects, assisting in handling goods, and even delivering materials to different parts of a facility [125]. The ability to function in various environments makes them integral to the concept of Industry 4.0, where human-robot collaboration is key to achieving flexible and scalable production. Given many of these platforms' cumbersome and heavy nature, safety and control are critical considerations. For this reason, many industrial porter robots are designed with a detachable control feature, allowing human operators to take manual control if necessary. This capability is handy in dynamic and unpredictable environments, where autonomous navigation might not always be reliable or safe. In such cases, the operator can switch to manual mode, guiding the robot through complex maneuvers or tasks that require human intuition and decision-making.

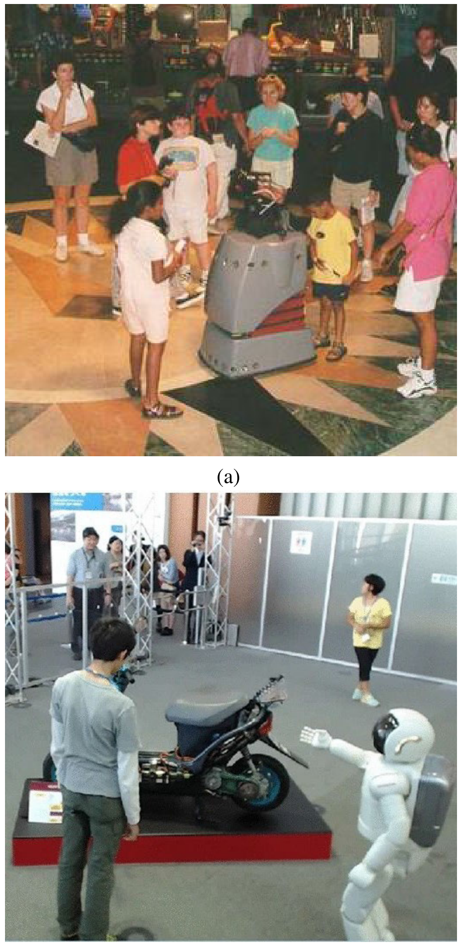
In the context of human-robot collaboration in warehouses, as in other applications, the line between following and guiding roles is often blurred. For instance, in [189] and [190], the authors examine the impact of different collaboration setups on productivity, accuracy, and human pick speed in order-picking tasks, comparing scenarios where the human leads the robot versus those where the robot leads. Their findings suggest that human-led configurations result in higher collaborative productivity, while robot-led configurations enhance accuracy. Additionally, they identify prevention regulatory focus as a key behavioral mechanism by which workers can increase their pick speed to bridge the productivity gap between these setups. This conclusion highlights how task dynamics and psychological factors play crucial roles in optimizing HRI. The study also demonstrates that manual order-picking skills are directly transferable to human-robot collaborative environments, emphasizing the importance of seamless integration between human expertise and robotic assistance. For effective collaboration, robots must adaptively switch between following and leading roles, anticipating human intentions and movements to maintain a high level of performance. This requirement involves sophisticated control algorithms capable of predicting human actions, enabling the robot to proactively adjust its behavior in real-time, whether by taking the lead or following the human's direction [52], [78]. Moreover, the flexibility of these systems is crucial in dynamic warehouse environments, where robots capable of dynamically switching between following and guiding roles allow for smoother transitions between tasks and improved overall efficiency.

C. TOUR GUIDES

Robotic tour guides represent some of the earliest and most impactful examples of autonomous robots interacting directly with the public, serving as pioneers in the field of human-aware robotics. These robots are designed not only to navigate complex, dynamic environments but also to engage and inform people, often focusing on enhancing the HRI experience over mere navigation. Their development and deployment have highlighted key advancements in autonomy, social interaction, and public-facing robotics. Two notable early examples of robotic tour guides are RHINO [164] and Minerva [204], both instrumental in laying the groundwork for future developments in this field.

RHINO, developed in the late 1990s, marked a significant breakthrough in the deployment of autonomous robots in public spaces. It was deployed in the Deutsches Museum in Bonn, Germany, where its primary task was to guide visitors through exhibitions while providing interactive tours (Figure 22(a)). A unique aspect of RHINO was its capability to allow users from around the world to control it remotely via a web interface, enabling them to set target locations and observe the robot in action. Building on RHINO's success, Minerva represented a significant evolution in robotic tour guides. Deployed at the Smithsonian's National Museum of American History in Washington, D.C., Minerva was equipped with more advanced HRI features and autonomous navigation capabilities. Unlike RHINO, Minerva utilized simultaneous localization and mapping (SLAM) techniques and ceiling mosaics for precise localization, allowing it to navigate autonomously in unfamiliar spaces.

The influence of RHINO and Minerva can be seen in subsequent robotic tour guides, which expanded on their foundational work. Many later systems integrated hand gesture recognition with voice feedback to improve the naturalness of HRI [22], [71]. These robots typically guide visitors through museums, offering information about exhibits through display screens or voice communication while adapting their behavior based on the situation. For example, positions of specific exhibitions can be pre-programmed or dynamically recognized using technologies like RFID tags or fiducial markers [158], [206]. Some robots go a step further by proactively approaching visitors who are observing exhibits [155], as illustrated in Figure 22(b), providing additional information while using expressive gestures, such as nodding, waving, or changing facial expressions, to capture attention and convey emotions [69], [193]. The ability to navigate crowded and cluttered environments is also crucial for these platforms, as museum spaces often present significant challenges in terms of dynamic obstacles and tight spaces [174].



(a)

(b)

FIGURE 22. Examples of typical robotic tour guides found in literature. Minerva, depicted in (a), was one of the first robots deployed as a museum tour guide [204]. In (b), Honda's ASIMO robot is presented in [205] as a receptionist for corporate environments, and employed in [155] as a tour guide.

Beyond museums, robotic guides are now used in a variety of environments, including universities, airports, and shopping malls. At universities, robots help new students and visitors navigate large campuses by providing information and guiding them to specific locations [207], [208]. In airports robots accompany travelers on connecting flights conveniently and efficiently from their arrival gate to the passport control [171]. In shopping malls, robots often lead customers to desired stores while maintaining a comfortable pace and delivering promotional messages or information about ongoing sales as they move through the space. After completing their guidance tasks, the robots typically return to predefined patrol routes or designated charging stations until they are needed again [209]. Another application of robotic guides is found in corporate environments, where receptionist robots greet visitors, guide them to meeting rooms or waiting areas, and notify relevant staff members of their arrival. These robots often integrate with online databases to manage visitor information efficiently, allowing for seamless check-ins and real-time updates on visitor movements throughout the building. As these platforms become more integrated with organizational systems, they contribute to more streamlined operations and improved visitor experiences [205].

IV. EVALUATION

This section critically evaluates the methods proposed in the literature, focusing on their feasibility and practicality based on the results and outcomes of various studies. The main metrics and evaluation modalities commonly used for each category are outlined, followed by a critical assessment of the most prominent and widely adopted state-of-the-art solutions.

As discussed in the following sections, perception and tracking systems are typically easier to evaluate due to the presence of well-established baselines and widely accepted metrics. This standardization simplifies the assessment of a system's effectiveness and reliability in person following and leading tasks, allowing researchers to quantitatively measure improvements and innovations with clarity and precision across different methodologies.

In contrast, evaluating navigation and social interaction aspects poses greater challenges due to their reliance on qualitative factors such as user experience, adaptability, and intuitive interaction. These aspects are inherently subjective and context-dependent, making it difficult to establish consistent and universally applicable evaluation criteria. While some quantitative metrics can assess certain aspects, they often miss factors of Human-Robot Interaction and the social dynamics essential for effective real-world integration. As a result, a rigorous evaluation approach is required, often combining quantitative measures with qualitative feedback. This methodology may include empirical testing in diverse settings, user studies to gather insights on interaction quality, and observational studies to assess real-world integration. Such a multifaceted approach yields a more comprehensive view of a system's effectiveness, acknowledging that success in controlled environments may not always extend to complex, real-world applications.

A. PERCEPTION AND TRACKING

In person following and guiding systems, perception and tracking are typically evaluated together, as they are often intrinsically linked, with perception informing the tracking process and vice versa. Accurate tracking relies heavily on the system's ability to perceive and identify the target individual consistently across time and space. Common performance metrics used to evaluate the quality of these systems include accuracy, precision, recall, and the F-score, which are computed across a sequence of consecutive frames [23], [76], [99], [108], [137]. These metrics provide valuable insights into the system's ability to consistently maintain recognition of the target and adjust its actions in response to changes in the environment or the person's movements. In addition to frame-based evaluations, certain systems assess performance by measuring the consistency of tracking over a specific following distance [34]. Furthermore, evaluating the ability to accurately localize the person during the following process is essential. In fact, identifying the target over time and maintaining a safe and effective following distance is critical in real-world applications, especially in dynamic environments [37], [145].

Predicting a person's future position has proven to be an essential component of high-performing systems, particularly in challenging environments with obstacles, rapid directional changes, or temporary occlusions. Systems that incorporate anticipatory tracking mechanisms demonstrate improved performance by ensuring continuous, uninterrupted tracking. Many recent approaches integrate predictive techniques such as Kalman filters, probabilistic tracking, or neural network-based methods [9], [52], [64], [77], [147]. These techniques allow the system to anticipate and respond to the target's future movements within a limited time horizon, reducing latency and enhancing real-time adaptability. Model based tracking techniques are even more effective, being able to predict not just immediate movements but also longer-term destinations [40], [127], [131], [132], [133], [134], [135]. Incorporating these predictive models not only increases the system's robustness but also helps mitigate potential disruptions caused by occlusion or sudden movement, further improving the overall reliability and effectiveness of person following and leading systems in practical applications [47], [56], [69].

Absolute localization systems such as UWB [152], [153], [154], multi-camera networks [35], [40], RFID systems [155], [156], [157], [158], and Angle of arrival technologies [77] represent some of the most accurate and robust tracking solutions currently available. These systems excel at providing high-precision localization in complex structured environments, often achieving sub-meter accuracy. For example, UWB is well-known for its ability to perform in cluttered indoor environments with minimal interference. At the same time, multi-camera networks offer wide-area coverage and redundancy through multiple viewpoints. RFID systems and AOA technologies provide similarly effective solutions, particularly in controlled environments like warehouses or laboratories, where they can leverage stationary sensors to achieve consistent tracking performance. However, as commented in Section II-B, these absolute localization systems come with a significant limitation: they rely heavily on structured environments. UWB, RFID, and AOA systems require fixed infrastructure such as anchors or antenna arrays, while multi-camera networks depend on the strategic placement of cameras to maintain visibility of the target. This dependence on environmental structuring poses challenges for deployment in unstructured settings, such as outdoor environments or dynamically changing spaces, where installing and maintaining infrastructure may

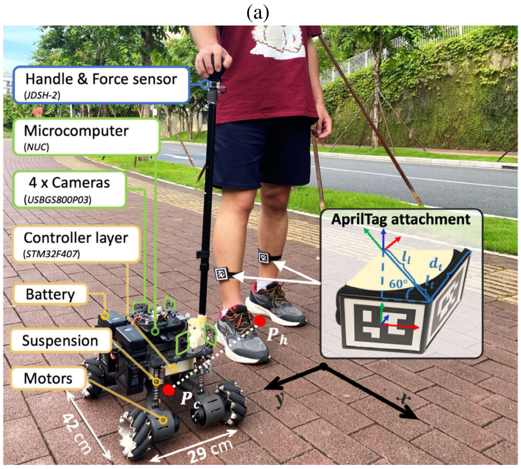
TABLE 5. Evaluation of perception and tracking methods. In the sixth column, a single tick indicates that the system can predict the next few immediate positions of the tracked person. In contrast, two ticks signify that the system demonstrates a deeper understanding of human intentions, enabling it to predict long-term movements or even the tracked person's final destination.

Detection method	Tracking method	Requires structuring environment	Requires structuring user	Structuring is inherent to application	Can predict future poses	Method is general	References
UWB	UWB	✓	✓	×	×	×	[152]–[154]
RFID	RFID	✓	✓	×	×	×	[149], [155]–[158]
Multi-camera marker recognition	Regression model	✓	✓	×	✓	×	[40]
Marker recognition	Frame-by-frame	×	✓	×	×	×	[4], [38], [41]–[45]
Ultrasonic, piezoelectric array	Frame-by-frame	×	✓	×	×	×	[149], [150]
Color based	Frame-by-frame	×	✓	×	×	×	[31], [33], [34], [36]
Physical contact	Physical contact	×	✓	✓	×	✓	[8], [79], [82], [117], [118], [134], [159], [159]
Tether contact	Tether following	×	✓	✓	×	✓	[75], [122], [160]–[162], [199]–[201]
Laser scan blobs geometric filtering	Kalman filters	×	×	-	✓	✓	[17], [18], [23]
DNNs	Kalman filters	×	×	-	✓	✓	[9], [10], [46], [47], [51], [52], [56], [64], [65], [67], [69], [147]
DNNs	Linear regressors	×	×	-	✓	✓	[50], [76], [99]
DNNs	Model based	×	×	-	✓✓	✓	[40], [127], [131]–[135]
DNNs	DNNs	×	×	-	✓✓	✓	[144], [145]

be impractical or prohibitively expensive. In many real-world scenarios, the necessity for fixed sensors limits the scalability and flexibility of these systems, making them unsuitable for applications where environmental modification is either undesirable or impossible, such as search-and-rescue operations or large-scale public environments.

Similarly, certain tracking methods require the person being followed to wear specific items that aid in recognition. This includes approaches where the target wears distinct clothing colors [31], [33], [34], [36], markers [4], [38], [39], [41], [42], [43], [44], [45], or devices such as ultrasonic, piezoelectric, or acceleration sensors [149], [150], [151]. Figure 23 presents some of these wearable devices. While these approaches avoid the need for extensive environmental modification, they introduce constraints on the user. These constraints can limit the applicability of such systems in situations where participants are unwilling or unable to wear the required equipment. In non-controlled environments, public spaces, or on-the-fly applications, the reliance on wearable elements can reduce the system's overall usability and acceptance. However, in scenarios where structuring the user is inherent to the application, such as healthcare, these approaches are both viable and effective. For instance, in medical applications where tethering between the robot and the user is necessary, such as in home oxygen therapy [160], [161] or with dog-like guide robots for

(a)



(b)

FIGURE 23. Illustrations of wearable markers used by perception systems for user recognition and tracking. In (a), infrared markers are embedded on a reflective jacket [43], while in (b), simple AprilTags are attached to the user's calves [45]. Such systems clearly impose constraints on the user, which may limit their applicability in scenarios where individuals are either unable or unwilling to wear the necessary equipment.

visually impaired users [75], [122], [162], [199], [200], [201], these methods offer a reliable way to maintain continuous interaction, ensuring that the robot remains in close proximity and responds to the user's movements. Other examples are scenarios where assistive robots may require patients to hold onto handles to ensure consistent tracking and avoid fall risks [8], [79], [82], [117], [118], [134], [159].

For any other use case, methods that are entirely robot-driven and do not rely on either environmental or user structuring are generally preferred. These autonomous systems offer greater flexibility and adaptability, as they can function independently of external infrastructure or user cooperation. Among these, laser-based approaches have proven to be highly efficient and reliable in person following tasks, offering precise distance measurements and obstacle detection [17], [18], [23], [29]. Laser-based sensors, such as LiDAR, are particularly effective in detecting objects with high accuracy, making them a popular choice for real-time tracking. However, these techniques are often part of more complex, multimodal systems that combine various sensing modalities for enhanced robustness and handle edge cases where a single sensor type may fail [20], [21], [22], [70], [87], [88]. Multimodal systems can combine data from laser, visual, and audio sensors to improve detection accuracy, handle occlusions, and deal with complex lighting or environmental conditions.

While laser-based systems have their strengths, visual methods leveraging advanced neural networks have emerged as the most effective and widely adopted solutions in recent years. These vision-based approaches outperform traditional methods in several key areas. They offer superior efficiency, lower computational overhead, and enhanced robustness under varying lighting conditions, making them adaptable to both indoor and outdoor environments [10], [46], [50], [51], [54], [55], [57], [65], [67], [73], [76], [99], [108], [109]. Deep learning-based visual systems have become the state-of-the-art for person following and tracking tasks due to their ability to generalize across diverse environments, adapt to complex scenarios, and handle occlusion and motion variability with remarkable precision. These methods are especially valuable in situations where the target may be partially obscured, move unpredictably, or where the lighting conditions fluctuate dramatically. One of the key advantages of neural network-based visual tracking systems is their ability to accurately predict the future long-term positions of the followed individual, or even their final destination, significantly improving tracking performance in dynamic settings [144], [145]. This predictive capability, alongside continuous advancements in hardware capabilities and neural network architectures, further enhances the real-time capabilities of these systems, making them increasingly viable for deployment in real-world applications.

To simplify the interpretation of the leading methodologies in perception and tracking for person following and guiding, Table 5 categorizes the key works discussed in this section. The categorization is based on previously introduced evaluation parameters, including the need for structured environments or user constraints, whether this structuring is intrinsically inherent to the task, and the ability to predict the tracked user's future positions. Finally, the generality of a method refers to its capability to be deployed on a robot and used in real-time, without requiring environmental or user-specific structuring, and without needing to fine-tune the system for individual users.

B. NAVIGATION AND INTERACTION

Unlike perception and tracking, evaluating navigation in person following and guiding tasks presents significant challenges. The absence of standardized datasets and benchmarks makes it difficult to achieve meaningful comparisons between the various methodologies proposed in the literature. While navigation can be assessed using well-established metrics—such as path length, smoothness, clearance time, mean velocities, and maximum accelerations—these traditional indicators do not always capture the complexities of social navigation. Person following and guiding fall within the realm of social robotics, where conventional navigation metrics may not fully reflect the system's effectiveness in real-world scenarios. For instance, a longer path with a higher

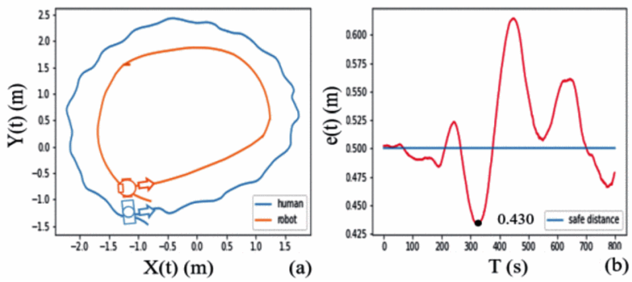


FIGURE 24. An example of a graphical representation for qualitative results, as shown in [153], allows researchers to visually compare paths, emphasizing the robot's capacity to follow human movement patterns. In this case, a side-by-side following scenario is examined, including measurements of the safety distance maintained by the robot relative to the person throughout the trial.

clearance time is not necessarily less optimal than a shorter, more direct route if it results in more socially acceptable behavior. A socially aware navigation system might prioritize leading the person on a safer, less disruptive route, even if it takes longer, or avoiding interactions that could interfere with other people in the environment. Therefore, traditional metrics must be interpreted carefully, considering the social context in which the robot operates.

To address the complexities of person following and leading systems, many studies incorporate qualitative evaluations to assess system performance. These qualitative approaches often involve graphical analyses of the trajectories followed by both the human and the robot during tasks, similar to that illustrated in Figure 24 [17], [18], [19], [68], [153], [182]. By visually comparing these paths, researchers can highlight the robot's ability to closely mirror or complement the human's movement patterns, offering insights into its tracking accuracy and behavioral alignment. While such graphical comparisons may not lend themselves easily to direct numerical comparisons across studies, they provide an intuitive, accessible interpretation of the system's spatial awareness, responsiveness, and overall functionality. These visual assessments are beneficial for demonstrating the robot's capacity to navigate dynamic environments and assume socially compliant behaviours, traits that are difficult to quantify but essential for real-world deployment. Another common form of qualitative evaluation is user studies, in which the proposed system is implemented on a real robotic platform and tested by a group of non-expert users. After each session, participants are typically asked to complete questionnaires providing subjective feedback on various aspects of their experience, such as the smoothness of the robot's movements, the naturalness of its interactions, and the perceived social acceptance of its behavior [15], [81], [83], [117], [118], [126], [197]. Despite the inherently subjective nature of this type of evaluation, user studies offer valuable insights, particularly when participants are drawn from the target audience of the system, whether they are visually impaired individuals, elderly, or museum visitors. By engaging the intended end-users, researchers can gather feedback that directly reflects the practical utility, user comfort, and real-world effectiveness of the system in fulfilling its designed purpose. User studies also help identify areas where the robot may need improvement in terms of human factors, such as how well it anticipates user movements, adjusts its speed and proximity, or reacts to unpredictable behaviors.

Some research has attempted to evaluate person following and leading systems using quantitative metrics. However, due to the lack of standardized evaluation settings and datasets, comparing results across different studies remains challenging, even when the same metrics are employed. Differences in the testing environments, robotic platforms, and participants involved introduce variability that hinders direct comparison between approaches. Furthermore, not all metrics used in these studies effectively capture the quality of social navigation. For example, some works have measured the robot's deviation from the precise human's path [77], [162], [200]. While this metric provides an indication of the robot's ability to follow, it has several limitations. First, it is not well-suited for dynamic environments where the path taken by the human may become obstructed over time, making it impractical or even unsafe for the robot to follow the exact same route. Second, forcing the robot to strictly adhere to the human's path can limit its autonomy and flexibility, preventing it from recognizing more efficient or socially acceptable alternative routes. As demonstrated in [15], allowing the robot to autonomously select its path toward the last known position of the human enhances adaptability and produces more human-like and socially compliant behaviours. This approach enables the robot to make context-aware decisions, balancing the need for proximity with the ability to navigate complex or crowded environments.

Other studies have evaluated system performance by conducting numerous trials and calculating the probability of successfully reaching the destination without colliding or engaging in socially inappropriate behaviors [30], [109], [185]. This probabilistic approach provides a robust measure of the system's reliability in real-world conditions and has been employed since the first social robotic studies with RHINO and Minerva [164]. Additionally, some works assess the mean distance a robot can travel before losing track of the user, which serves as a crucial indicator of the navigation system's effectiveness in supporting the tracking module, particularly in dynamic or complex environments [15]. This metric highlights the system's ability to maintain consistent performance over extended periods of time and distance, ensuring stable user engagement. Another common metric used in literature is the distance maintained between the robot and the followed person during the task [78], [168]. These tests demonstrate how the system can to respond to changes of direction and speed assumed by the followed person and react accordingly. However, also in this case these types of evaluations often struggle to serve as robust, reusable benchmarks for researchers and roboticists worldwide. This problem arises because they inherently depend on the specific

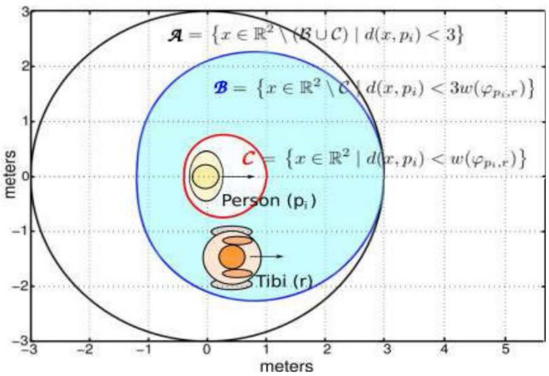


FIGURE 25. Diagram of areas used in [133] for the evaluation of the robot's performance. The area $\mathcal{C}$ represents the person's personal space, $\mathcal{A}$ represent a socially acceptable area, while $\mathcal{B}$ is the optimal area where the robot should remain during the person accompaniment.

testing environment, platforms used, and the particular user being followed or guided, significantly limiting the reproducibility and generalization of the evaluation methods across different contexts.

In recent years, driven by the rapid growth of social robotics research, many new social evaluation methods have emerged. Most of these primarily focus on defining quantitative metrics based on people's personal space and comfortableness criteria [136]. For example, in [131], the authors introduced distinct spatial zones. They define a socially acceptable area where the robot can remain without causing discomfort, an optimal area for the robot to stay in, that is both socially acceptable and conducive to interaction—allowing the user to easily see and engage with the robot—and a personal space area for each individual in the environment, which the robot should avoid crossing. Figure 25 shows a graphical representation of these evaluation areas. Performances are then evaluated by measuring how much time the robot spends in the socially acceptable and unacceptable areas during the accompaniment task. These evaluations are frequently complemented by user studies, which provide a more subjective and human-centered perspective. Combining quantitative metrics with user feedback offers a more holistic understanding of how effectively the robot navigates human environments, assessing not only technical accuracy but also social integration and comfort [12], [133], [178].

From an efficiency and feasibility point of view, while traditional navigation planners and controllers are still largely employed for human following and guiding tasks [45], [154], [157], they often fall short in accounting for the complexities inherent in crowded, human-populated environments. These conventional systems typically lack the flexibility to handle dynamic interactions, spontaneous human behavior, or the subtleties of social norms, making them less suitable for real-world social navigation. As a result, there is a growing preference for methods specifically designed to address social navigation challenges more effectively. One of the most widely adopted models in this domain is the Social Force Model, which excels at balancing social acceptance with goal-oriented navigation. SFM incorporates human-centric factors such as personal space and social comfort while predicting future human positions, even in dense environments [40], [127], [132], [135]. This predictive ability to account for human intentions and movement trajectories makes it particularly valuable in scenarios where robots need to coexist with unpredictable human behaviors. In parallel, as in the broader field of robotic navigation, Deep Reinforcement Learning (DRL) has emerged as a highly promising approach for human following and guiding tasks. Reinforcement Learning is a machine learning paradigm where an agent learns an optimal policy from repeated experiences. It involves exploring an environment and receiving feedback for each action performed through rewards or penalties. The agent updates its strategy over time to maximize the cumulative reward and improve performance. DRL methods can autonomously learn from interactions within human environments, including simulated scenarios, to optimize navigation strategies. DRL agents specifically designed for human following and guiding have demonstrated superior performance compared to traditional approaches in understanding social factors, predicting human movement, and avoiding collisions in dynamic, complex environments [171], [182], [185]. However, as later discussed in V-B, reinforcement learning training typically requires large amounts of experience. Creating realistic simulation environments and valid reward functions is usually challenging, resulting in an overall difficulty in training effective agents. This problem often limits the applicability of reinforcement learning techniques in favor of other more robust and consistent solutions.

V. OPEN PROBLEMS AND FUTURE DIRECTIONS

This section highlights several active research areas and underexplored topics in literature. While these areas present significant challenges, they hold great potential and could offer valuable advancements in the field of person following and guiding.

A. METRICS FOR SOCIAL NAVIGATION AND INTERACTION

As discussed in Section IV-B, social factors are difficult to evaluate with quantitative metrics. The actual common practice to benchmark new methods is through qualitative user studies, where participants are asked to compile questionnaires providing subjective feedback on various aspects of their testing experience [210]. Typical metrics used in interviews to score social navigation are smoothness, friendliness, unobtrusiveness, and avoidance foresight [211]. On the other hand, some quantitative metrics have been used to evaluate navigation in recent works: proxemics and social spaces intrusion to measure the respect of intimate distances, and social work to measure the impact of robot's motion on other agents [212], [213]. Specific metrics for the human following task should also aim at inspecting the capacity of the robot to track and not lose the target person, as highlighted in [9] with a relative heading metric. New metrics could be conceived by mixing social acceptance scores with proxemics and quantitative analysis of trajectories and commands. Nonetheless, an in-depth investigation on the relevant and missing aspects evaluated by the existing metrics is an open field.

Besides the definition of metrics, effective benchmarking is another crucial and unsolved problem in following, guiding and, more broadly, social navigation tasks. While several simulation environments for algorithm evaluation have been realized [212], [214], real-world testing is still an open issue. A key obstacle is the strict requirement to collect ground truth trajectories of both humans and robots in order to evaluate the robot motion quantitatively. The development of a standardized methodology for collecting and evaluating such data would greatly benefit the scientific community, enabling consistent and fair comparisons across studies, encouraging reproducibility, and driving faster progress in the field of robotics. A promising approach researchers could prioritize involves leveraging multi-camera absolute localization systems [35], [40]. Although expensive and requiring an initial precise calibration, these systems offer unmatched accuracy in tracking multiple targets simultaneously.

B. LEARNING FROM EXPERIENCE

With the rise of advanced neural networks and machine learning technologies, novel methods have also emerged in the field of person following and guiding. While machine learning has been extensively applied in perception and tracking modules, Deep Reinforcement Learning has shown significant promise, particularly for end-to-end navigation [182], [183], [184], [185] and motion control [147]. DRL allows robots to learn complex behaviors through trial and error, making it a powerful tool for autonomous decision-making. However, despite its potential, there has been limited exploration of Inverse Reinforcement Learning (IRL) [171], [215], and even fewer works have focused on Imitation Learning (IL) for these tasks.

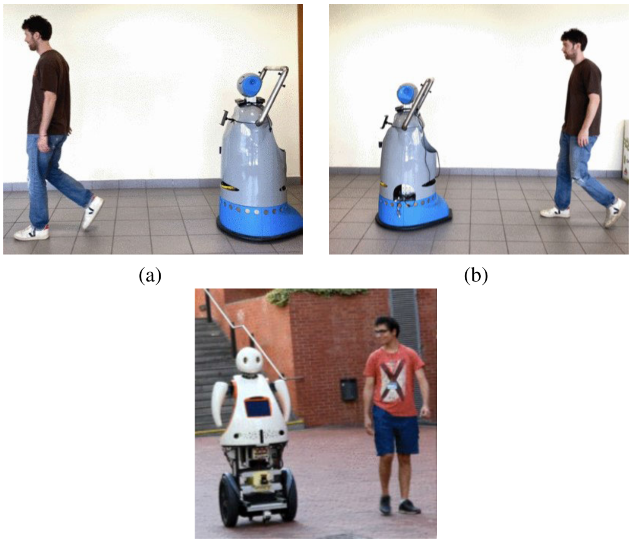
Reinforcement learning (RL) training typically demands large amounts of experience to develop effective agents. Collecting such data in real-world environments can be prohibitive due to safety risks, time constraints, and the complexity of the task. As a result, researchers often rely on sophisticated simulation environments to train RL agents. However, for tasks as intricate as person following and guiding, creating accurate and realistic simulations is particularly challenging. The gap between simulation and real-world environments often leads to significant performance degradation when agents trained in simulations are deployed in the real world. This discrepancy can result from environmental variations or unmodeled dynamics that are difficult to fully capture in simulations. In contrast, Imitation Learning offers a promising alternative by leveraging demonstrations from expert policies, such as human actions. IL can potentially capture subtle human behaviors and social aspects that are difficult to model in simulations. This ability to mimic expert behavior can allow robots to perform more naturally in human-centered environments. However, the potential of IL in person following and guiding remains underexplored, and more research is needed to fully realize its advantages in real-world applications.

Future research should also focus on hybrid approaches that combine the strengths of RL and IL, incorporating human demonstrations with reinforcement learning techniques to overcome the challenges posed by both simulation limitations and real-world complexities. Additionally, advancements in transfer learning, domain adaptation, and sim-to-real techniques can help bridge the gap between simulated and physical environments, enabling more robust and efficient person following and guiding systems.

With the latest advancements, researchers are trying to develop general intelligence models learning from huge amounts of data. These models, commonly known as Foundation Models, have been inspired by the success raised in Natural Language Processing, and are now entering the robotics world. The main idea is to learn a general representation of the data associated with an entire set of tasks. A task-specific model can then be obtained with little fine-tuning. However, learning a general robot foundation model is a tough objective [216], requiring tons of data. Collaborative manipulation tasks already adopt this kind of solution [217], [218]. They usually involve confined spaces and objects, in contrast with the wide range of environments encountered in autonomous navigation. Therefore, studying more general and abstract representations for observation data instead of environment-specific features and raw sensor measurements could lead to a more effective generalization in unseen conditions.

C. LARGE MODELS FOR HRI

Large Language Models (LLM) and Large Visual-Language Models (LVLM) represent the latest successful achievement of Deep Learning in the robotics field. Starting from the inception of ChatGPT [219], Large Language Models have demonstrated powerful capabilities worldwide. These models trained on huge amounts of text, images, or both types of data can drastically enhance communication and understanding capabilities of a social robot [220]. Recently, open-source versions of some LVLM have been released, paving the way for wide adoption of these solutions in research. The introduction of LVLM contextually opens diverse challenges [221]. First, these large models are required to run on robotic embedded hardware without drastically increasing power consumption. Offline processing of visual-audio data avoids typical privacy issues from exposing private and domestic data to the cloud. Second, the optimal way to prompt these models is still under strong investigation. At the same time, coupling the perception and control module of a robot with this powerful interface could lead to a broad simplification of user commands, thanks to the higher context interpretation offered by large models. Another potential



(a) (b) (c)

FIGURE 26. Three different examples of robot navigation configuration. In (a) and (b), respectively person following and guiding, as presented in [222]. In (c), side-by-side accompaniment presented in [223].

role for VLM is tied to the long process required to learn complex behaviors from experience. VLMs can be employed to substitute human feedback or quantitative reward design, leveraging their general task understanding to improve the learning process.

D. TO GUIDE OR TO FOLLOW?

Human following and guiding tasks share numerous similarities, and the distinction between them is often blurred. Both tasks involve detecting and tracking the human, planning and executing a socially acceptable path, and frequently engaging in interaction with the person throughout the process. In both scenarios, the robot must navigate dynamic environments while considering the human's movements, intentions, and social aspects. Additionally, both tasks require real-time adaptability to ensure the robot behaves in ways that are safe, predictable, and respectful of personal space. Figure 26 represents the three main navigation configurations: the robot follows the human, guides them, and accompanies them side-by-side. A growing body of research suggests that robotic systems could greatly benefit from moving beyond the traditional leader-follower paradigm, where one agent consistently leads and the other follows. Instead, adopting a more dynamic, peer-to-peer interaction model, where roles can shift fluidly based on context, environmental conditions, or task demands—could enhance the flexibility and efficiency of human-robot collaboration [130]. Despite the potential benefits of hybrid systems that combine both following and guiding behaviors, there is still relatively little research focused on directly comparing the effectiveness of these two approaches joined in a single, adaptive system. [140], [188]. Most studies in literature tend to treat following and guiding as separate tasks, evaluating them independently rather than exploring how they might work together in a more fluid, switchable system [189], [190], [196], [222]. Evaluating the effectiveness of hybrid follower-leader systems could provide valuable insights for more flexible and capable robots. Further research into hybrid methods could also explore the underlying algorithms needed to enable smooth transitions between leading and following behaviors. Moreover, the evaluation of these systems should not focus solely on technical metrics. HRI studies could investigate how users perceive and respond to dynamic role-switching behaviors, examining factors such as trust, comfort, and perceived safety.

E. FOLLOW AHEAD

In recent years, the follow-ahead methodology has gained increasing attention, driven by advancements in novel and innovative technologies [145]. This approach is particularly beneficial in specific scenarios where it is advantageous for the robot to position itself ahead of the user while still performing a following, rather than a guiding, task. For example, head-free shopping carts [51] follow users in malls but are positioned in front for quick and easy access, which is especially useful for elderly or physically impaired customers who may require immediate assistance. Similarly, cane-robots are designed to move ahead of users, providing balance and mobility support [8], [45]. Another relevant application is autonomous luggage systems, where following ahead can give the user an improved sense of security and awareness by keeping the robot within their line of sight.

However, the follow-ahead approach presents more significant challenges than traditional following from behind, which explains the scarcity of research on this topic in literature. In this configuration, the robot must not only track the user's movements but also anticipate their intentions to maintain an optimal position ahead. This necessity requires real-time predictions of the user's direction, speed, and potential sudden adjustments while ensuring smooth, non-intrusive navigation. Moreover, the robot must also quickly recover from incorrect predictions or actions and adapt to unexpected changes in the user's behavior or surrounding environment. To address these challenges, future research should focus on developing more intelligent systems capable of effectively predicting human intentions and promptly recovering from erroneous decisions. As discussed in Section V-C, the enhanced contextual understanding provided by Large Visual-Language Models offers a promising direction for advancing this field, enabling robots to interpret complex human behaviors and environmental conditions with greater accuracy.

F. FOLLOW A GROUP

People often form groups to complete daily tasks more efficiently, which means that robots must develop the ability to interact not only with individuals but also with groups. One critical aspect of this interaction is the robot's ability to follow or accompany a group of people toward a designated location. The first challenge lies in defining what constitutes a "group," as this concept can vary depending on the context. Early research on robotic guidance, such as [171] and [172], defined a group as successfully following as long as a single member stayed behind the robot. Other approaches employed multiple robots, referred to as "shepherds," to assist in keeping the group together, ensuring that all members followed the designated leader robot [140]. These "shepherds" actively help maintain group cohesion while the leading robot directs the overall movement. More recent studies have adopted a side-by-side accompaniment strategy, allowing robots to better detect and monitor each individual within the group. This approach enables the robot to switch dynamically between leader and follower roles, depending on the situation [178], [223]. By staying beside the group, the robot can more accurately perceive the group's dynamics and adapt to changes.

Once a group is identified, the next significant challenge is modeling its behavior. While the group could be simplified as a single entity, in practice, it exhibits complex spatial interactions and social behaviors [224]. Incorrect modeling can complicate the robot's ability to predict future movements and trajectories, leading to errors in navigation and coordination. Future research should prioritize the development of more sophisticated and realistic models of group behavior. Such advancements could enable the application of existing, highly efficient person following algorithms to group following tasks without requiring entirely new, dedicated algorithms. Moreover, these enhanced models could facilitate the creation of navigation systems capable of smoothly transitioning between person following and group following modes, eliminating the need for additional, specialized solutions.

G. REPRESENTATIVE AND LONG-TERM TESTING

In evaluating person following and guiding systems, simulation environments often fail to fully capture the complexities of real-world scenarios. Consequently, many studies opt for real-world experimentation involving human subjects to test the systems in more authentic settings. However, the test samples used in these evaluations do not always represent the target population. For instance, while some studies aimed at elderly users successfully conducted tests in retirement homes and assisted living facilities [79], [197], or tested smart robotic canes with visually impaired individuals [80], [83], others had to rely on healthy participants as stand-ins for their target demographic [8], [45], [74], [75], [122], [159]. This substitution diminishes the effectiveness and relevance of the evaluations, as testing on healthy subjects can yield significantly different results from those that might be observed with the intended user base. Testing on the actual end-users, especially those with disabilities, could reveal unique challenges, usability concerns, and insights that are crucial for refining these systems.

Moreover, a notable gap in the current literature is the absence of long-term testing of these technologies. This lack is often due to practical challenges, such as the inability of facilities to host researchers for extended periods or the difficulty in securing patient participation for long and continuous trials. However, long-term evaluations are critical for assessing the durability, reliability, and user adaptability of person following systems over time. Research in this area should encourage collaborations with care facilities and directly involve target users to enable meaningful, long-term deployments and evaluations. Such partnerships would provide opportunities to test person-following and guiding systems in authentic settings over extended periods, uncovering insights into their durability, reliability, and adaptability.

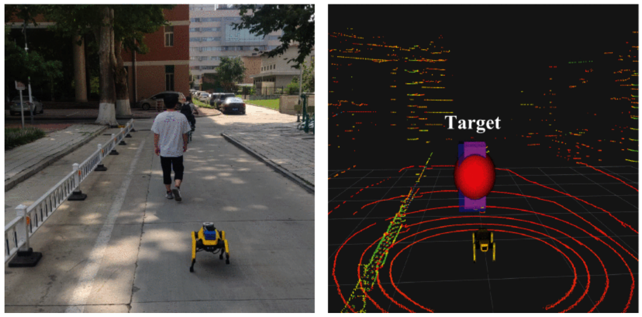


FIGURE 27. Example of outdoor person following with a quadrupedal robot [96]. The figure on the right shows the sparse point cloud of a 3D LiDAR mounted on the platform.

H. OUTDOOR PERSON FOLLOWING

Since its inception, person-following and guiding robots have primarily been deployed within indoor environments. This preference arises from the fact that most applications requiring robotic assistance, as discussed in Section III, are conducted indoors. Key application domains include healthcare and logistics, where tasks are typically performed in enclosed spaces such as hospitals, living facilities, warehouses, and homes. Similarly, tour guide robots are commonly used in controlled environments like museums, universities, airports, and shopping malls. The predominance of indoor use is also linked to the significant challenges posed by unstructured outdoor environments. Outdoor settings often feature uneven and rough terrain, presenting obstacles for wheeled robots designed for flat surfaces. To address these challenges, the limited research on outdoor person-following robots has explored alternative platforms, including aerial robots like drones [38] and legged robots [54], [96], [121]. Other approaches restrict operation to structured paths, such as sidewalks or roads [45]. Additionally, outdoor environments often lack detailed maps, and traditional localization and planning methods may struggle to ensure smooth and efficient navigation in such scenarios. For this reason, more advanced and accurate perception sensors are often employed, as shown in Figure 27.

Despite these challenges, advancing research in this area is essential to develop systems capable of maintaining reliable user following functionality during transitions between indoor and outdoor settings. These solutions have the potential to unlock unprecedented applications for robotic assistance, such as search and rescue missions, exploration tasks, and disaster response operations, where robots are often still teleoperated. Multi-modal solutions can significantly enhance the localization and navigation capabilities of autonomous robots, particularly during transitions between indoor and outdoor environments. Integrating data from multiple sensors can leverage the strengths of each modality to compensate for the limitations of others. For instance, Global Navigation Satellite Systems (GNSS) provide robust outdoor localization but are unreliable indoors, whereas visual and LiDAR-based systems excel in structured indoor settings but struggle in dynamic or unstructured outdoor environments. Additionally, multi-modal approaches can use machine learning to fuse sensor data contextually, improving the robustness of navigation systems in complex and diverse real-world scenarios.

I. HIGH-SPEED FOLLOWING

While the existing literature on robotic following systems has primarily focused on applications in healthcare, logistics, and tour guiding, where robots are required to adapt to slower human speeds—often significantly below their maximum capabilities—there has been relatively little research addressing the challenge of following at high speeds. Aerial robots, such as drones, are exceptionally well-suited for high-speed following tasks due to their agility and ability to navigate complex environments from a vantage point that is less restricted by ground-level obstacles [38]. High-speed following presents unique advantages in various fields, including search and rescue operations, high-performance rehabilitation, and sports training, where the ability to track and assist fast-moving individuals is crucial.

However, the high-speed following task introduces a range of complexities not encountered in slower applications. The system must adapt to faster dynamics in several ways. First, the tracking module needs to operate at a much higher frame rate to reliably capture the user's rapid movements and positions. This requires both faster sensors and more efficient algorithms to process data in real-time. Additionally, the robot must make immediate, sometimes unpredictable navigation decisions to keep up with the fast-moving person, avoiding obstacles while maintaining safe and effective trajectories. Achieving high velocities is another critical challenge. The robot must not only accelerate quickly to catch up with the person but also sustain these speeds without compromising stability or safety. At higher speeds, ensuring that the robot remains close to the user becomes more difficult. Small errors in tracking or path planning can in fact result in the robot falling behind or veering off course. Moreover, high-speed operation increases the demand for responsive control systems. The controller has to adjust the robot's motion dynamically based on rapid changes in the user's pace, direction, or environmental factors. In addition to the technical hurdles, high-speed following systems must also account for safety concerns. The faster the robot moves, the greater the potential for accidents, especially in cluttered or unpredictable environments. Therefore, the design of such systems must include robust fail-safes and adaptive behaviors to avoid collisions and ensure user safety even at elevated speeds.

Given these challenges, developing robots capable of high-speed following remains an emerging area that demands further exploration and innovation. Researchers should focus on advancing faster, more responsive systems capable of handling the rapid dynamics and complex decision-making required for such tasks. As discussed in V-B, reinforcement learning techniques offer a promising direction for addressing these challenges. These techniques have demonstrated impressive capabilities in achieving precise and robust control, even under extreme conditions, such as those encountered in high-speed drone racing [225].

VI. CONCLUSION

In this survey, we have presented a comprehensive review of the current state-of-the-art in the field of person following and guiding systems for autonomous mobile robots. Our analysis has explored the critical components of these systems, focusing on key areas such as perception, tracking, navigation, and human-robot interaction. We also highlighted the evolution of these technologies and their deployment across a wide range of application fields, including healthcare, personal assistance, logistics, and tour guiding. By examining these approaches, we have emphasized both the technological advancements and their practical applications in real-world environments. One of the key contributions of this survey is the critical evaluation of the various approaches, assessing their feasibility, practicality, and performance as reported in the literature. Furthermore, we have outlined the open challenges that persist in the development of person following and guiding robots. These challenges represent critical avenues for future research and innovation, and overcoming them is key to unlocking the full potential of these systems in practical, real-world applications.

We hope this work serves as a valuable resource for researchers seeking to navigate the broad field of person following and guiding, providing insights into current trends and helping to identify future challenges. Future research should focus on addressing the current limitations, particularly in terms of real-world deployment, system adaptability, and coherent interaction in dynamic environments. By tackling these challenges, researchers can unlock the full potential of person following and guiding robots, enabling a new era of collaborative, socially intelligent robotics in everyday life.

LIST OF ACRONYMS

AKP	Anticipative Kinodynamic Planner.15, 16
AOA	Angle of arrival.12, 26
ASFD	Adaptive-Safe-Following Distance.15
ASFM	Aerial Social Force Model.16
BFS	Breadth First Search. 13
CNN	Convolutional Neural Network. 5, 7, 8, 11
CTMap	Conditional Transition Maps.14
DNN	Deep Neural Network.10,27
DRL	Deep Reinforcement Learning.16,30,31

DWA	Dynamic Window Approach. 14-16
EKF	Extended Kalman Filter. 9
FMM-DWA	Fast Marching Method Dynamic Window Approach. 15
GDIF	Generic Distance-Invariant Features. 8
GNN	Global Nearest Neighbor. 10
HCSN	Human-Centered Sensitive Navigation.15
HOG	Histogram of Oriented Gradients. 5, 6, 11
HRI	Human-Robot Interaction. 1, 14, 17, 24-26, 32
HTSD	Highest-Traversability-Score Direction. 15
IMU	Inertial Measurement Unit. 8
JPDAF	Joint Probabilistic Data Association Filter. 10
KLT	Kanade-Lucas-Tomasi. 11
LiDAR	Light Detection and Ranging. 3-5, 7, 8, 28, 33, 34
LRF	Laser Range Finder. 3-5, 7, 10, 12
MPC	Model Predictive Control. 14, 15
MPEPC	Model Predictive Equilibrium Point Control. 15
PID	Proportional-Integral-Derivative. 14, 15
PRM	Probabilistic Roadmap. 13
ProMPs	Probabilistic Movement Primitives. 9
RABS	Robust-Adaptive-Behavior Strategy. 15
RBFSVM	Radial Basis Function Support Vector Machine. 8
RFID	Radio Frequency Identification. 12, 25-27
RGB	Red-Green-Blue. 7, 9, 11
RGBD	Red-Green-Blue-Depth. 5, 8, 11
RLS-PAA	Recursive Least Square Parameter Adaptation Algorithm. 15
RRT	Rapidly-exploring Random Trees. 13, 14
SFM	Social Force Model. 16, 30
SORT	Simple Online and Realtime Tracking. 10
SSD	Single Shot Detector. 7
SVDD	Support Vector Data Description. 5, 8
SVM	Support Vector Machine. 5, 7, 8
T-CTMap	Temporal Conditional Transition Maps. 14, 15
TOF	Time-of-flight. 3, 4
TTS	text-to-speech. 18, 19
UKF	Unscented Kalman Filter. 9
UWB	Ultra-Wideband. 12, 26, 27
YOLO	You Only Look Once. 5, 7

REFERENCES

[1] M. Sitti, "Physical intelligence as a new paradigm," *Extreme Mech. Lett.*, vol. 46, Jul. 2021, Art. no. 101340.

[2] Y. Li, Z. Li, Y. Duan, and A.-B. Spulber, "Physical artificial intelligence (PAI): The next-generation artificial intelligence," *Frontiers Inf. Technol. Electron. Eng.*, vol. 24, no. 8, pp. 1231–1238, Aug. 2023.

[3] R. Tasaki, M. Kitazaki, J. Miura, and K. Terashima, "Prototype design of medical round supporting robot 'Terapio,'" in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Sep. 2015, pp. 829–834.

[4] S. Mahajan and C. M. Vidhyapathi, "Design of a medical assistant robot," in *Proc. 2nd IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol. (RTEICT)*, May 2017, pp. 877–881.

[5] H. Yoo, D. Kim, J. Sohn, K. Lee, and C. Kim, "Development of a worker-following robot system: Worker position estimation and motion control under measurement uncertainty," *Machines*, vol. 11, no. 3, p. 366, Mar. 2023.

[6] M. M. Islam, A. Lam, H. Fukuda, Y. Kobayashi, and Y. Kuno, "A person-following shopping support robot based on human pose skeleton data and LiDAR sensor," in *Proc. 15th Int. Conf. Intell. Comput. Methodologies (ICIC)*, Nanchang, China. Cham, Switzerland: Springer, Jan. 2019, pp. 9–19.

[7] M. M. Islam, A. Lam, H. Fukuda, Y. Kobayashi, and Y. Kuno, "Person-following shopping support robot using Kinect depth camera based on 3D skeleton tracking," in *Proc. 16th Int. Conf. Intell. Comput. Methodologies (ICIC)*, Bari, Italy. Cham, Switzerland: Springer, Jan. 2020, pp. 28–37.

[8] Q. Yan, J. Huang, Z. Yang, Y. Hasegawa, and T. Fukuda, "Human-following control of cane-type walking-aid robot within fixed relative posture," *IEEE/ASME Trans. Mechatronics*, vol. 27, no. 1, pp. 537–548, Feb. 2022.

[9] A. Eirale, M. Martini, and M. Chiaberge, "Human-centered navigation and person-following with omnidirectional robot for indoor assistance and monitoring," *Robotics*, vol. 11, no. 5, p. 108, Oct. 2022.

[10] A. Eirale, M. Martini, L. Tagliavini, D. Gandini, M. Chiaberge, and G. Quaglia, "Marvin: An innovative omni-directional robotic assistant for domestic environments," *Sensors*, vol. 22, no. 14, p. 5261, Jul. 2022.

[11] (2017). *Piaggio-Fast-Forward (2017) Gita*. Accessed: Oct. 27, 2024. [Online]. Available: https://piaggiofastforward.com/

[12] E. Repiso, A. Garrell, and A. Sanfeliu, "Adaptive side-by-side social robot navigation to approach and interact with people," *Int. J. Social Robot.*, vol. 12, no. 4, pp. 909–930, Aug. 2020.

[13] S. S. Honig, T. Oron-Gilad, H. Zaichyk, V. Sarne-Fleischmann, S. Olatunji, and Y. Edan, "Toward socially aware person-following robots," *IEEE Trans. Cognit. Develop. Syst.*, vol. 10, no. 4, pp. 936–954, Dec. 2018.

[14] M. J. Islam, J. Hong, and J. Sattar, "Person-following by autonomous robots: A categorical overview," *Int. J. Robot. Res.*, vol. 38, no. 14, pp. 1581–1618, Dec. 2019.

[15] R. Gockley, J. Forlizzi, and R. Simmons, "Natural person-following behavior for social robots," in *Proc. 2nd ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Mar. 2007, pp. 17–24.

[16] N. Kawarazaki, L. T. Kuwae, and T. Yoshidome, "Development of human following mobile robot system using laser range scanner," *Proc. Comput. Sci.*, vol. 76, pp. 455–460, Apr. 2015.

[17] J. Yuan, S. Zhang, Q. Sun, G. Liu, and J. Cai, "Laser-based intersection-aware human following with a mobile robot in indoor environments," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 354–369, Oct. 2018.

[18] T. V. Nguyen, M. H. Do, and J. Jo, "Robust-adaptive-behavior strategy for human-following robots in unknown environments based on fuzzy inference mechanism," *Ind. Robot, Int. J. Robot. Res. Appl.*, vol. 49, no. 6, pp. 1089–1100, Sep. 2022.

[19] N. V. Toan, M. D. Hoang, P. B. Khoi, and S.-Y. Yi, "The human-following strategy for mobile robots in mixed environments," *Robot. Auto. Syst.*, vol. 160, Feb. 2023, Art. no. 104317.

[20] M. Kobilarov, G. Sukhatme, J. Hyams, and P. Batavia, "People tracking and following with mobile robot using an omnidirectional camera and a laser," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Jun. 2006, pp. 557–562.

[21] A. Cosgun, D. A. Florencio, and H. I. Christensen, "Autonomous person following for telepresence robots," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2013, pp. 4335–4342.

[22] V. Alvarez-Santos, R. Iglesias, X. M. Pardo, C. V. Regueiro, and A. Canedo-Rodriguez, "Gesture-based interaction with voice feedback for a tour-guide robot," *J. Vis. Commun. Image Represent.*, vol. 25, no. 2, pp. 499–509, Feb. 2014.

[23] A. Leigh, J. Pineau, N. Olmedo, and H. Zhang, "Person tracking and following with 2D laser scanners," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2015, pp. 726–733.

[24] U. Uguzlar, E. Cansu, E. C. Contarli, and V. Sezer, "Autonomous human following robot based on follow the gap method," in *Proc. IEEE Int. Conf. Auto. Robot Syst. Competitions (ICARSC)*, Apr. 2023, pp. 139–144.

[25] Y. Kobayashi, Y. Kinpara, T. Shibusawa, and Y. Kuno, "Robotic wheelchair based on observations of people using integrated sensors," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2009, pp. 2013–2018.

[26] Y. Sato, M. Arai, R. Suzuki, Y. Kobayashi, Y. Kuno, K. Yamazaki, and A. Yamazaki, "A maneuverable robotic wheelchair able to move adaptively with a caregiver by considering the situation," in *Proc. IEEE RO-MAN*, Aug. 2013, pp. 282–287.

[27] W. Chung, H. Kim, Y. Yoo, C.-B. Moon, and J. Park, "The detection and following of human legs through inductive approaches for a mobile robot with a single laser range finder," *IEEE Trans. Ind. Electron.*, vol. 59, no. 8, pp. 3156–3166, Aug. 2012.

[28] Y. Sung and W. Chung, "Hierarchical sample-based joint probabilistic data association filter for following human legs using a mobile robot in a cluttered environment," *IEEE Trans. Hum.-Mach. Syst.*, vol. 46, no. 3, pp. 340–349, Jun. 2016.

[29] D. Cha and W. Chung, "Human-leg detection in 3D feature space for a person-following mobile robot using 2D LiDARs," *Int. J. Precis. Eng. Manuf.*, vol. 21, no. 7, pp. 1299–1307, Jul. 2020.

[30] H. Yao, H. Dai, E. Zhao, P. Liu, and R. Zhao, "Laser-based Side-by-Side following for human-following robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 2651–2656.

[31] C. Schlegel, J. Illmann, H. Jaberg, M. Schuster, and R. Wörz, "Vision based person tracking with a mobile robot," in *Proc. Brit. Mach. Vis. Conf.*, 1998, pp. 1–10.

[32] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer, "Multi-camera multi-person tracking for EasyLiving," in *Proc. 3rd IEEE Int. Workshop Vis. Surveill.*, Aug. 2000, pp. 3–10.

[33] H. Kwon, Y. Yoon, J. Byung Park, and A. C. Kak, "Person tracking with a mobile robot using two uncalibrated independently moving cameras," in *Proc. IEEE Int. Conf. Robot. Autom.*, Sep. 2005, pp. 2877–2883.

[34] S. Waldherr, R. Romero, and S. Thrun, "A gesture based interface for human–robot interaction," *Auto. Robots*, vol. 9, pp. 151–173, Sep. 2000.

[35] K. Morioka, J.-H. Lee, and H. Hashimoto, "Human-following mobile robot in a distributed intelligent sensor network," *IEEE Trans. Ind. Electron.*, vol. 51, no. 1, pp. 229–237, Feb. 2004.

[36] H.-M. Gross, H. Boehme, C. Schroeter, S. Mueller, A. Koenig, E. Einhorn, C. Martin, M. Merten, and A. Bley, "TOOMAS: Interactive shopping guide robots in everyday use–final implementation and experiences from long-term field trials," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2009, pp. 2005–2012.

[37] J. D. Brookshire, "Person following using histograms of oriented gradients," U.S. Patent 12 848 677, Feb. 3, 2011.

[38] E. Graether and F. Mueller, "Joggobot: A flying robot as jogging companion," in *Proc. CHI Extended Abstr. Human Factors Comput. Syst.*, May 2012, pp. 1063–1066.

[39] Z. Wu and S. Payandeh, "Toward design of a drip-stand patient follower robot," *J. Robot.*, vol. 2020, pp. 1–16, Mar. 2020.

[40] A. Garrell, L. Garza-Elizondo, M. Villamizar, F. Herrero, and A. Sanfeliu, "Aerial social force model: A new framework to accompany people using autonomous flying robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 7011–7017.

[41] M. F. Ahmad, S. S. N. Alhady, S. Kaharuddin, and W. A. F. W. Othman, "Visual based sensor cart follower for wheelchair by using microcontroller," in *Proc. IEEE Int. Conf. Control Syst., Comput. Eng. (ICCSCE)*, Nov. 2015, pp. 123–128.

[42] K. Kanchanasatian, "A robot companion algorithm for Side-by-Side object tracking and following," in *Proc. 37th Int. Tech. Conf. Circuits/Syst., Comput. Commun. (ITC-CSCC)*, Jul. 2022, pp. 1–7.

[43] M. T. K. Tsun, B. T. Lau, and H. S. Jo, "An improved indoor robot human-following navigation model using depth camera, active IR marker and proximity sensors fusion," *Robotics*, vol. 7, no. 1, p. 4, Jan. 2018.

[44] K. Jeyatharan, H. E. M. H. B. Ekanayake, and K. D. Sandaruwan, "Behavior based tracking for human following robot," in *Proc. IEEE 16th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Dec. 2021, pp. 371–376.

[45] H. Liu, F. Wu, B. Zhong, Y. Zhao, J. Zhang, W. Niu, and M. Zhang, "Close-range human following control on a cane-type robot with multi-camera fusion," *IEEE Robot. Autom. Lett.*, vol. 8, no. 10, pp. 6443–6450, Oct. 2023.

[46] H. Masuzawa, J. Miura, and S. Oishi, "Development of a mobile robot for harvest support in greenhouse horticulture — Person following and mapping," in *Proc. IEEE/SICE Int. Symp. Syst. Integr. (SII)*, Dec. 2017, pp. 541–546.

[47] L. Pang, Z. Cao, J. Yu, P. Guan, X. Chen, and W. Zhang, "A robust visual person-following approach for mobile robots in disturbing environments," *IEEE Syst. J.*, vol. 14, no. 2, pp. 2965–2968, Jun. 2020.

[48] I. Condés, J. Fernández-Conde, E. Perdices, and J. M. Cañas, "Robust person identification and following in a mobile robot based on deep learning and optical tracking," *Electronics*, vol. 12, no. 21, p. 4424, Oct. 2023.

[49] S. Banerjee, A. Kumar, and A. Shekhar, "Indoor surveillance robot with person following and re-identification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2024, pp. 1–8.

[50] B.-J. Lee, J. Choi, C. Baek, and B.-T. Zhang, "Robust human following by deep Bayesian trajectory prediction for home service robots," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 7189–7195.

[51] P. Nikdel, R. Shrestha, and R. Vaughan, "The hands-free push-cart: Autonomous following in front by predicting user trajectory around obstacles," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 4548–4554.

[52] H. Wu, W. Xu, B. Yao, Y. Hu, and H. Feng, "Interacting multiple model-based adaptive trajectory prediction for anticipative human following of mobile industrial robot," *Proc. Comput. Sci.*, vol. 176, pp. 3692–3701, Jan. 2020.

[53] A. Boschi, F. Salvetti, V. Mazzia, and M. Chiaberge, "A cost-effective person-following system for assistive unmanned vehicles with deep learning at the edge," *Machines*, vol. 8, no. 3, p. 49, Aug. 2020.

[54] Y. Zhao, Y. Gao, Q. Sun, Y. Tian, L. Mao, and F. Gao, "A real-time low-computation cost human-following framework in outdoor environment for legged robots," *Robot. Auto. Syst.*, vol. 146, Dec. 2021, Art. no. 103899.

[55] R. K. Megalingam, R. Anandu, D. HemaTejaAnirudhBabu, G. Sriram, and V. S. YashwanthAvvari, "Implementation of a person following robot in ROS-gazebo platform," in *Proc. Int. Conf. Advancement Technol. (ICONAT)*, Jan. 2022, pp. 1–5.

[56] H. Ye, J. Zhao, Y. Pan, W. Cherr, L. He, and H. Zhang, "Robot person following under partial occlusion," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2023, pp. 7591–7597.

[57] A. A. F. Sakri, I. M. Lazim, S. A.-T. Mauzi, M. Sahrim, L. Ramli, and A. Noordin, "Autonomous person-following telepresence robot using monocular camera and deep learning YOLO," *Appl. Model. Simul.*, vol. 8, pp. 101–109, Aug. 2024.

[58] F. Rollo, A. Zunino, G. Raiola, F. Amadio, A. Ajoudani, and N. Tsagarakis, "FollowMe: A robust person following framework based on visual re-identification and gestures," in *Proc. IEEE Int. Conf. Adv. Robot. Its Social Impacts (ARSO)*, Jun. 2023, pp. 84–89.

[59] Y. Li, J. Zhao, and X. Luo, "SLAM-based person following robot with online trajectory recovery strategy," in *Proc. 35th Chin. Control Decis. Conf. (CCDC)*, May 2023, pp. 4645–4650.

[60] H. Zhang and X. Ren, "Robust real-time object tracking system for human-following quadruped robot," in *Proc. Chin. Intell. Syst. Conf.* Cham, Switzerland: Springer, Jan. 2022, pp. 388–397.

[61] B. X. Chen, R. Sahdev, and J. K. Tsotsos, "Integrating stereo vision with a CNN tracker for a person-following robot," in *11th Int. Conf. Proc. Comput. Vis. Syst. (ICVS)*, Shenzhen, China. Cham, Switzerland: Springer, Jan. 2017, pp. 300–313.

[62] Md. J. Islam, M. Fulton, and J. Sattar, "Toward a generic diver-following algorithm: Balancing robustness and efficiency in deep visual detection," *IEEE Robot. Autom. Lett.*, vol. 4, no. 1, pp. 113–120, Jan. 2019.

[63] Z. Rui, W. Zhaokui, and Z. Yulin, "A person-following nanosatellite for in-cabin astronaut assistance: System design and deep-learning-based astronaut visual tracking implementation," *Acta Astronautica*, vol. 162, pp. 121–134, Sep. 2019.

[64] K. Koide, J. Miura, and E. Menegatti, "Monocular person tracking and identification with on-line deep feature selection for person following robots," *Robot. Auto. Syst.*, vol. 124, Feb. 2020, Art. no. 103348.

[65] J. Liu, X. Chen, C. Wang, G. Zhang, and R. Song, "A person-following method based on monocular camera for quadruped robots," *Biomimetic Intell. Robot.*, vol. 2, no. 3, Sep. 2022, Art. no. 100058.

[66] J. Laplaza, N. Rodríguez, J. E. Domínguez-Vidal, F. Herrero, S. Hernández, A. López, A. Sanfeliu, and A. Garrell, "IVO robot: A new social robot for human–robot collaboration," in *Proc. 17th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Mar. 2022, pp. 860–864.

[67] J. Montesdeoca, J. M. Toibero, J. Jordan, A. Zell, and R. Carelli, "Person-following controller with socially acceptable robot motion," *Robot. Auto. Syst.*, vol. 153, Jul. 2022, Art. no. 104075.

[68] N. Yao, E. Anaya, Q. Tao, S. Cho, H. Zheng, and F. Zhang, "Monocular vision-based human following on miniature robotic blimp," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 3244–3249.

[69] M. Bennewitz, F. A. Faber, D. Joho, M. Schreiber, and S. Behnke, "Towards a humanoid museum guide robot that interacts with multiple persons," in *Proc. 5th IEEE-RAS Int. Conf. Humanoid Robots*, Jan. 2006, pp. 418–423.

[70] J. Fritsch, M. Kleinehagenbrock, S. Lang, G. A. Fink, and G. Sagerer, "Audiovisual person tracking with a mobile robot," in *Proc. Int. Conf. Intell. Auto. Syst.*, Jan. 2004, pp. 898–906.

[71] A. Clodic, S. Fleury, R. Alami, R. Chatila, G. Bailly, L. Brethes, M. Cottret, P. Danes, X. Dollat, F. Elisei, I. Ferrane, M. Herrb, G. Infantes, C. Lemaire, F. Lerasle, J. Manhes, P. Marcoul, P. Menezes, and V. Montreuil, "Rackham: An interactive robot-guide," in *Proc. 15th IEEE Int. Symp. Robot Hum. Interact. Commun. (ROMAN)*, Sep. 2006, pp. 502–509.

[72] N. Van Toan, S.-H. Bach, and S.-Y. Yi, "A hierarchical approach for updating targeted person states in human-following mobile robots," *Intell. Service Robot.*, vol. 16, no. 3, pp. 287–306, Jul. 2023.

[73] A. Tomoya, S. Nakayama, A. Hoshina, and M. Sugaya, "A mobile robot for following, watching and detecting falls for elderly care," *Proc. Comput. Sci.*, vol. 112, pp. 1994–2003, Aug. 2017.

[74] H. Wakabayashi, Y. Hiroi, K. Miyawaki, and A. Ito, "Development of a personal guide robot that leads a guest Hand-in-Hand while keeping a distance," *Sensors*, vol. 24, no. 7, p. 2345, Apr. 2024.

[75] G. Galatas, C. Mcmurrough, G. L. Mariottini, and F. Makedon, "EyeDog: An assistive-guide robot for the visually impaired," in *Proc. 4th Int. Conf. Pervasive Technol. Rel. Assistive Environ.*, May 2011, pp. 1–8.

[76] R. Algabri and M.-T. Choi, "Deep-learning-based indoor human following of mobile robot using color feature," *Sensors*, vol. 20, no. 9, p. 2699, May 2020.

[77] D. Jin, Z. Fang, and J. Zeng, "A robust autonomous following method for mobile robots in dynamic environments," *IEEE Access*, vol. 8, pp. 150311–150325, 2020.

[78] J. E. Sierra-García, V. Fernández-Rodríguez, M. Santos, and E. Quevedo, "Development and experimental validation of control algorithm for person-following autonomous robots," *Electronics*, vol. 12, no. 9, p. 2077, Apr. 2023.

[79] S. Dubowsky, F. Genot, S. Godding, H. Kozono, A. Skwersky, H. Yu, and L. S. Yu, "Pamm–a robotic aid to the elderly for mobility assistance and monitoring: A," in *Proc. IEEE Int. Conf. Robot. Autom. Symposia (ICRA) Millennium*, vol. 1, Sep. 2000, pp. 570–576.

[80] V. Kulyukin, C. Gharpure, and J. Nicholson, "RoboCart: Toward robot-assisted navigation of grocery stores by the visually impaired," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2005, pp. 2845–2850.

[81] T. Yoshimi, M. Nishiyama, T. Sonoura, H. Nakamoto, S. Tokura, H. Sato, F. Ozaki, N. Matsuhira, and H. Mizoguchi, "Development of a person following robot with vision based target detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2006, pp. 5286–5291.

[82] A. Morris, R. Donamukkala, A. Kapuria, A. Steinfeld, J. T. Matthews, J. Dunbar-Jacob, and S. Thrun, "A robotic Walker that provides guidance," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 1, Nov. 2003, pp. 25–30.

[83] J. Guerreiro, D. Sato, S. Asakawa, H. Dong, K. M. Kitani, and C. Asakawa, "CaBot: Designing and evaluating an autonomous navigation robot for blind people," in *Proc. 21st Int. ACM SIGACCESS Conf. Comput. Accessibility*, Oct. 2019, pp. 68–82.

[84] A. Azarbayejani and A. Pentland, "Real-time self-calibrating stereo person tracking using 3-D shape estimation from blob features," in *Proc. 13th Int. Conf. Pattern Recognit.*, vol. 3, 1996, pp. 627–632.

[85] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, Jul. 1997.

[86] T. Tsai and C. Yao, "A robust tracking algorithm for a human-following mobile robot," *IET Image Process.*, vol. 15, no. 3, pp. 786–796, Feb. 2021.

[87] M. Eisenbach, A. Vorndran, S. Sorge, and H.-M. Gross, "User recognition for guiding and following people with a mobile robot in a clinical environment," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 3600–3607.

[88] L. Susperregi, J. M. Martínez-Otzeta, A. Ansuategui, A. Ibarguren, and B. Sierra, "RGB-D, laser and thermal sensor fusion for people following in a mobile robot," *Int. J. Adv. Robotic Syst.*, vol. 10, no. 6, p. 271, Jun. 2013.

[89] T. Inoue, Y. Okazaki, and K. Itoya, "Person following algorithm with pixel-area addition method of thermal sensors for autonomous mobile robots," in *Proc. 10th Int. Conf. Control, Autom. Robot. (ICCAR)*, Apr. 2024, pp. 77–82.

[90] T. Inoue and Y. Okazaki, "Human-following control of omnidirectional autonomous mobile robot by integrating sensor information," in *Proc. 23rd Int. Conf. Control, Autom. Syst. (ICCAS)*, Oct. 2023, pp. 1055–1059.

[91] C. C. Ulloa, D. Orbea, J. del Cerro, and A. Barrientos, "A tracking-guidance people method using a quadruped robot and artificial vision," in *Proc. 10th Int. Conf. Autom., Robot. Appl. (ICARA)*, Feb. 2024, pp. 196–200.

[92] J. Sales, R. Marín, E. Cervera, S. Rodríguez, and J. Pérez, "Multi-sensor person following in low-visibility scenarios," *Sensors*, vol. 10, no. 12, pp. 10953–10966, Dec. 2010.

[93] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR05)*, vol. 1, Sep. 2005, pp. 886–893.

[94] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Feb. 2001. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/990517

[95] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.

[96] X. Chen, J. Liu, J. Wu, C. Wang, and R. Song, "LoPF: An online LiDAR-only person-following framework," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022.

[97] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, Netherlands. Cham, Switzerland: Springer, Jan. 2016, pp. 21–37.

[98] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[99] R. Algabri and M.-T. Choi, "Target recovery for robust deep learning-based person following in mobile robots: Online trajectory prediction," *Appl. Sci.*, vol. 11, no. 9, p. 4165, May 2021.

[100] S. Manzoor, E.-J. Kim, S.-H. Bae, and T.-Y. Kuc, "Edge deployment of vision-based model for human following robot," in *Proc. 23rd Int. Conf. Control, Autom. Syst. (ICCAS)*, Oct. 2023, pp. 1721–1726.

[101] R. Algabri and M.-T. Choi, "Online boosting-based target identification among similar appearance for person-following robots," *Sensors*, vol. 22, no. 21, p. 8422, Nov. 2022.

[102] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[103] N. T. Binh, B. N. Dung, L. X. Chieu, N. Long, M. Soklin, N. D. Thanh, H. X. Tung, N. V. Dung, N. D. Truong, and L. M. Hoang, "Deep learning-based object tracking and following for agv robot," in *Proc. Int. Conf. Intell. Syst. Netw.* Cham, Switzerland: Springer, 2023, pp. 204–214.

[104] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[105] I. Asante, L. B. Theng, M. T. K. Tsun, H. S. Jo, and C. McCarthy, "Segmentation-based angular position estimation algorithm for dynamic path planning by a person-following robot," *IEEE Access*, vol. 11, pp. 41034–41053, 2023.

[106] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[107] B. X. Chen, R. Sahdev, and J. K. Tsotsos, "Person following robot using selected online ada-boosting with stereo camera," in *Proc. 14th Conf. Comput. Robot Vis. (CRV)*, May 2017, pp. 48–55.

[108] W. Chi, J. Wang, and M. Q.-H. Meng, "A gait recognition method for human following in service robots," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 9, pp. 1429–1440, Sep. 2018.

[109] A. Jevtic, G. Doisy, Y. Parmet, and Y. Edan, "Comparison of interaction modalities for mobile indoor robot guidance: Direct physical interaction, person following, and pointing control," *IEEE Trans. Hum.-Mach. Syst.*, vol. 45, no. 6, pp. 653–663, Dec. 2015.

[110] A. Wang, Y. Makino, and H. Shinoda, "Machine learning-based human-following system: Following the predicted position of a walking human," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 4502–4508.

[111] A. Kendall, M. Grimes, and R. Cipolla, "PoseNet: A convolutional network for real-time 6-DOF camera relocalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2938–2946.

[112] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7291–7299.

[113] R. Votel and N. Li. (2021). *Next-Generation Pose Detection With MoveNet and TensorFlow.js*. Accessed: Oct. 27, 2024. [Online]. Available: https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html

[114] S. Kean, J. C. Hall, P. Perry, S. Kean, J. C. Hall, and P. Perry, "Microsoft's Kinect SDK," in *Meet the Kinect: An Introduction To Programming Natural User Interfaces*. Berkeley, CA, USA: Apress, 2011, pp. 151–173.

[115] K. Koide and J. Miura, "Identification of a specific person using color, height, and gait features for a person following robot," *Robot. Auto. Syst.*, vol. 84, pp. 76–87, Oct. 2016.

[116] C. Weinrich, T. Wengefeld, C. Schroeter, and H.-M. Gross, "People detection and distinction of their walking aids in 2D laser range data based on generic distance-invariant features," in *Proc. 23rd IEEE Int. Symp. Robot Human Interact. Commun.*, Aug. 2014, pp. 767–773.

[117] M. E. Pollack, L. E. Brown, D. Colbry, C. Orosz, B. Peintner, S. Ramakrishnan, S. Engberg, J. T. Matthews, J. Dunbar-Jacob, C. E. McCarthy, S. Thrun, M. Montemerlo, J. Pineau, and N. Roy, "Pearl: A mobile robotic assistant for the elderly," in *Proc. AAAI Workshop Autom. Eldercare*, Edmonton, AB, Canada, Jan. 2002, pp. 85–91.

[118] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun, "Towards robotic assistants in nursing homes: Challenges and results," *Robot. Auto. Syst.*, vol. 42, nos. 3–4, pp. 271–281, Mar. 2003.

[119] E. Chebotareva, R. Safin, K.-H. Hsia, A. Carballo, and E. Magid, "Person-following algorithm based on laser range finder and monocular camera data fusion for a wheeled autonomous mobile robot," in *Proc. Int. Conf. Interact. Collaborative Robot.* Cham, Switzerland: Springer, Jan. 2020, pp. 21–33.

[120] E. Chebotareva, K.-H. Hsia, K. Yakovlev, and E. Magid, "Laser rangefinder and monocular camera data fusion for human-following algorithm by PMB-2 mobile robot in simulated gazebo environment," in *Proc. 15th Int. Conf. Electromechanics Robot. 'Zavalishin's Readings' ER(ZR)*, Ufa, Russia. Cham, Switzerland: Springer, Sep. 2020, pp. 357–369.

[121] S. Siva and H. Zhang, "Robot perceptual adaptation to environment changes for long-term human teammate following," *Int. J. Robot. Res.*, vol. 41, no. 7, pp. 706–720, Jun. 2022.

[122] S. Saegusa, Y. Yasuda, Y. Uratani, E. Tanaka, T. Makino, and J.-Y. Chang, "Development of a guide-dog robot: Human–robot interface considering walking conditions for a visually handicapped person," *Microsyst. Technol.*, vol. 17, nos. 5–7, pp. 1169–1174, Jun. 2011.

[123] N. Hirose, R. Tajima, and K. Sukigara, "Personal robot assisting transportation to support active human life," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 5395–5402.

[124] J. Chen and W.-J. Kim, "A human-following mobile robot providing natural and universal interfaces for control with wireless electronic devices," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 5, pp. 2377–2385, Oct. 2019.

[125] E. Petrović, A. Leu, D. Ristić-Durrant, and V. Nikolić, "Stereo vision-based human tracking for robotic follower," *Int. J. Adv. Robotic Syst.*, vol. 10, no. 5, p. 230, May 2013.

[126] Y. Morales, T. Kanda, and N. Hagita, "Walking together: Side-by-side walking model for an interacting robot," *J. Hum.-Robot Interact.*, vol. 3, no. 2, pp. 50–73, 2014.

[127] E. Repiso, A. Garrell, and A. Sanfeliu, "Robot approaching and engaging people in a human–robot companion framework," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 8200–8205.

[128] D. Karunarathne, Y. Morales, T. Kanda, and H. Ishiguro, "Model of side-by-side walking without the robot knowing the goal," *Int. J. Social Robot.*, vol. 10, no. 4, pp. 401–420, Sep. 2018.

[129] V. T. Nguyen, T. Duc Tran, and I.-H. Kuo, "A model for determining natural pathways for side-by-side companion robots in passing pedestrian flows using dynamic density," in *Proc. 31st IEEE Int. Conf. Robot Hum. Interact. Commun. (RO-MAN)*, Aug. 2022, pp. 1107–1114.

[130] V. T. Nguyen, C. Jayawardena, and I. Ardekani, "A navigation model for side-by-side robotic wheelchairs for optimizing social comfort in crossing situations," *Robot. Auto. Syst.*, vol. 100, pp. 27–40, Feb. 2018.

[131] G. Ferrer, A. G. Zulueta, F. H. Cotarelo, and A. Sanfeliu, "Robot social-aware navigation framework to accompany people walking side-by-side," *Auto. Robots*, vol. 41, no. 4, pp. 775–793, Apr. 2017.

[132] A. Cosgun, E. A. Sisbot, and H. I. Christensen, "Anticipatory robot path planning in human environments," in *Proc. 25th IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN)*, Aug. 2016, pp. 562–569.

[133] G. Ferrer, A. Garrell, and A. Sanfeliu, "Robot companion: A social-force based approach with human awareness-navigation in crowded environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 1688–1694.

[134] A. Nanavati, X. Z. Tan, J. Connolly, and A. Steinfeld, "Follow the robot: Modeling coupled human–robot dyads during navigation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 3836–3843.

[135] E. Repiso, G. Ferrer, and A. Sanfeliu, "On-line adaptive side-by-side human robot companion in dynamic urban environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 872–877.

[136] S. Sekiguchi, A. Yorozu, K. Kuno, M. Okada, Y. Watanabe, and M. Takahashi, "Uncertainty-aware non-linear model predictive control for human-following companion robot," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 8316–8322.

[137] M. Gupta, S. Kumar, L. Behera, and V. K. Subramanian, "A novel vision-based tracking algorithm for a human-following mobile robot," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 7, pp. 1415–1427, Jul. 2017.

[138] X. Xing, W. Li, S. Yuan, and Y. Li, "Fuzzy logic-based arbitration for shared control in continuous Human–Robot collaboration," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 7, pp. 3979–3991, Jul. 2024.

[139] M. Wang, Y. Liu, D. Su, Y. Liao, L. Shi, J. Xu, and J. Valls Miro, "Accurate and real-time 3-D tracking for the following robots by fusing vision and ultrasonar information," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 3, pp. 997–1006, Jun. 2018.

[140] A. Garrell and A. Sanfeliu, "Cooperative social robots to accompany groups of people," *Int. J. Robot. Res.*, vol. 31, no. 13, pp. 1675–1701, Nov. 2012.

[141] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3645–3649.

[142] H. Ye, J. Zhao, Y. Zhan, W. Chen, L. He, and H. Zhang, "Person re-identification for robot person following with online continual learning," *IEEE Robot. Autom. Lett.*, vol. 9, no. 11, pp. 9151–9158, Nov. 2024.

[143] M. A. Syed, Y. Ou, T. Li, and G. Jiang, "Lightweight multimodal domain generic person reidentification metric for person-following robots," *Sensors*, vol. 23, no. 2, p. 813, Jan. 2023.

[144] Z. Wu, M. Hashimoto, B. Guo, and K. Takahashi, "A path prediction method for human-accompanying mobile robot based on neural network," in *Proc. 2nd Sino-Foreign-Interchange Workshop Intell. Sci. Intell. Data Eng. (IScIDE)*, Xi'an, China. Cham, Switzerland: Springer, 2011, pp. 35–42.

[145] M. Mahdavian, P. Nikdel, M. TaherAhmadi, and M. Chen, "STPOTR: Simultaneous human trajectory and pose prediction using a non-autoregressive transformer for robot follow-ahead," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2023, pp. 9959–9965.

[146] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3464–3468.

[147] A. Eirale, M. Martini, and M. Chiaberge, "RL-DWA omnidirectional motion planning for person following in domestic assistance and monitoring," in *Proc. 9th Int. Conf. Autom., Robot. Appl. (ICARA)*, Feb. 2023, pp. 86–90.

[148] T. Motokucho and N. Oda, "Vision-based human-following control using optical flow field for power assisted wheelchair," in *Proc. IEEE 13th Int. Workshop Adv. Motion Control (AMC)*, Mar. 2014, pp. 266–271.

[149] J. Sales, J. V. Martí, R. Marín, E. Cervera, and P. J. Sanz, "CompaRob: The shopping cart assistance robot," *Int. J. Distrib. Sensor Netw.*, vol. 12, no. 2, Feb. 2016, Art. no. 4781280.

[150] T. Pingali, E. Lemaire, and N. Baddour, "Ultrasonic tethering to enable side-by-side following for powered wheelchairs," *Sensors*, vol. 19, no. 1, p. 109, Dec. 2018.

[151] R. Tasaki, H. Sakurai, and K. Terashima, "Moving target localization method using foot mounted acceleration sensor for autonomous following robot," in *Proc. IEEE Conf. Control Technol. Appl. (CCTA)*, Aug. 2017, pp. 827–833.

[152] J. Sarmento, F. N. dos Santos, A. S. Aguiar, V. Filipe, and A. Valente, "Fusion of time-of-flight based sensors with monocular cameras for a robotic person follower," *J. Intell. Robotic Syst.*, vol. 110, no. 1, p. 30, Mar. 2024.

[153] G. Xue, H. Yao, Y. Zhang, J. Huang, L. Zhu, and H. Dai, "UWB-based adaptable side-by-side following for human-following robots," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2022, pp. 333–338.

[154] C. V. Dang, H. Ahn, J.-W. Kim, and S. C. Lee, "Collision-free navigation in human-following task using a cognitive robotic system on differential drive vehicles," *IEEE Trans. Cognit. Develop. Syst.*, vol. 15, no. 1, pp. 78–87, Mar. 2023.
[155] T. Iio, S. Satake, T. Kanda, K. Hayashi, F. Ferreri, and N. Hagita, "Human-like guide robot that proactively explains exhibits," *Int. J. Social Robot.*, vol. 12, no. 2, pp. 549–566, May 2020.
[156] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, "An affective guide robot in a shopping mall," in *Proc. 4th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Mar. 2009, pp. 173–180.
[157] C. Wu, B. Tao, H. Wu, Z. Gong, and Z. Yin, "A UHF RFID-based dynamic object following method for a mobile robot using phase difference information," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.
[158] B. P. E. Alvarado Vásquez and F. Matía, "A tour-guide robot: Moving towards interaction with humans," *Eng. Appl. Artif. Intell.*, vol. 88, Feb. 2020, Art. no. 103356.
[159] D. M. Stramel, R. M. Carrera, S. A. Rahok, J. Stein, and S. K. Agrawal, "Effects of a person-following light-touch device during overground walking with visual perturbations in a virtual reality environment," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 4139–4146, Oct. 2019.
[160] A. Tani, G. Endo, E. F. Fukushima, S. Hirose, M. Iribe, and T. Takubo, "Study on a practical robotic follower to support home oxygen therapy patients-development and control of a mobile platform," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2011, pp. 2423–2429.
[161] G. Endo, B. Allan, Y. Iemura, E. F. Fukushima, M. Iribe, T. Takubo, and M. Ohira, "Mobile follower robot as an assistive device for home oxygen therapy—Evaluation of tether control algorithms," *ROBOMECH J.*, vol. 2, no. 1, pp. 1–9, Dec. 2015.
[162] H.-K. Wu, P.-Y. Chen, H.-Y. Wu, and C.-H. Yu, "User local coordinate-based accompanying robot for human natural movement of daily life," *Sensors*, vol. 21, no. 11, p. 3889, Jun. 2021.
[163] N. Chen, X. Chen, C. Chen, Y. Leng, and C. Fu, "Research on the human-following method, fall gesture recognition, and protection method for the walking-aid cane robot," in *Proc. IEEE Int. Conf. Cyborg Bionic Syst. (CBS)*, Mar. 2023, pp. 286–291.
[164] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, "The museum tour-guide robot RHINO," *Autonome Mobile Systeme*, vol. 1999, pp. 245–254, Jan. 1999.
[165] J. Peng, Z. Liao, H. Yao, Z. Su, Y. Zeng, and H. Dai, "MPC-based human-accompanying control strategy for improving the motion coordination between the target person and the robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2023, pp. 7969–7975.
[166] W. Lei, R. Wang, T. Liang, and Q. Ren, "An autonomous robot for collision-free person following through model predictive control," in *Proc. 49th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2023, pp. 1–6.
[167] A. Kumar Ashe and K. M. Krishna, "Maneuvering intersections & occlusions using MPC-based prioritized tracking for differential drive person following robot," in *Proc. IEEE 17th Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2021, pp. 1352–1357.
[168] M. Samadi Gharajeh and H. B. Jond, "Speed control for leader-follower robot formation using fuzzy system and supervised machine learning," *Sensors*, vol. 21, no. 10, p. 3433, May 2021.
[169] G. A. Acosta-Amaya, D. A. Miranda-Montoya, and J. A. Jimenez-Builes, "Lightweight two-layer control architecture for human-following robot," *Sensors*, vol. 24, no. 23, p. 7796, Dec. 2024.
[170] G. Ferrer and A. Sanfeliu, "Proactive kinodynamic planning using the extended social force model and human motion prediction in urban environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2014, pp. 1730–1735.
[171] R. Triebel et al., "SPENCER: A socially aware service robot for passenger guidance and help in busy airports," in *Proc. Results 10th Int. Conf. Field Service Robot.* Cham, Switzerland: Springer, Jan. 2016, pp. 607–622.
[172] M. Joosse and V. Evers, "A guide robot at the airport: First impressions," in *Proc. Companion ACM/IEEE Int. Conf. Hum.-Robot Interact.*, Mar. 2017, pp. 149–150.
[173] T. Kucner, J. Saarinen, M. Magnusson, and A. J. Lilienthal, "Conditional transition maps: Learning motion patterns in dynamic environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 1196–1201.
[174] R. Philippsen and R. Siegwart, "Smooth and efficient obstacle avoidance for a tour guide robot," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 1, Oct. 2003, pp. 446–451.
[175] J. J. Park and B. Kuipers, "Autonomous person pacing and following with model predictive equilibrium point control," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2013, pp. 1060–1067.
[176] C.-P. Lam, C.-T. Chou, K.-H. Chiang, and L.-C. Fu, "Human-centered robot navigation—towards a harmoniously human–robot coexisting environment," *IEEE Trans. Robot.*, vol. 27, no. 1, pp. 99–112, Feb. 2011.
[177] G. Ferrer and A. Sanfeliu, "Anticipative kinodynamic planning: Multi-objective robot navigation in urban and dynamic environments," *Auto. Robots*, vol. 43, no. 6, pp. 1473–1488, Aug. 2019.
[178] E. Repiso, F. Zanlungo, T. Kanda, A. Garrell, and A. Sanfeliu, "People's V-formation and side-by-side model adapted to accompany groups of people by social robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 2082–2088.
[179] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 51, no. 5, pp. 4282–4286, May 1995.
[180] E. Repiso, A. Garrell, and A. Sanfeliu, "Adaptive social planner to accompany people in real-life dynamic environments," *Int. J. Social Robot.*, vol. 16, no. 6, pp. 1189–1221, Jun. 2024.
[181] H. Shin and S.-E. Yoon, "Optimization-based path planning for person following using following field," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 11352–11359.
[182] L. Pang, Y. Zhang, S. Coleman, and H. Cao, "Efficient hybrid-supervised deep reinforcement learning for person following robot," *J. Intell. Robotic Syst.*, vol. 97, no. 2, pp. 299–312, Feb. 2020.
[183] C. K. Dewa and J. Miura, "Integrating multiple policies for person-following robot training using deep reinforcement learning," *IEEE Access*, vol. 9, pp. 75526–75541, 2021.
[184] B. S. B. Dewantara and J. Miura, "Generation of a socially aware behavior of a guide robot using reinforcement learning," in *Proc. Int. Electron. Symp. (IES)*, Sep. 2016, pp. 105–110.
[185] L. Kästner, B. Fatloun, Z. Shen, D. Gawrisch, and J. Lambrecht, "Human-following and -guiding in crowded environments using semantic deep-reinforcement-learning for mobile service robots," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2022, pp. 833–839.
[186] H. S. Shanee, K. Dror, O.-G. Tal, and E. Yael, "The influence of following angle on performance metrics of a human-following robot," in *Proc. 25th IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN)*, Aug. 2016, pp. 593–598.
[187] S. Olatunji, T. Oron-Gilad, V. Sarne-Fleischmann, and Y. Edan, "User-centered feedback design in person-following robots for older adults," *J. Paladyn Behav. Robot.*, vol. 11, no. 1, pp. 86–103, Apr. 2020.
[188] A. Noormohammadi-Asl, S. L. Smith, and K. Dautenhahn, "To lead or to follow? Adaptive robot task planning in human–robot collaboration," 2024, *arXiv:2401.01483*.
[189] A. Pasparakis, J. De Vries, and M. De Koster, "In control or under control? Human–robot collaboration in warehouse order picking," *Hum.-Robot Collaboration Warehouse Order Picking*, 2021. [Online]. Available: https://doi.org/10.2139/ssrn.3816533
[190] A. Pasparakis, J. De Vries, and R. De Koster, "Assessing the impact of human–robot collaborative order picking systems on warehouse workers," *Int. J. Prod. Res.*, vol. 61, no. 22, pp. 7776–7790, Nov. 2023.
[191] A. Cherubini, R. Passama, B. Navarro, M. Sorour, A. Khelloufi, O. Mazhar, S. Tarbouriech, J. Zhu, O. Tempier, A. Crosnier, P. Fraisse, and S. Ramdani, "A collaborative robot for the factory of the future: BAZAR," *Int. J. Adv. Manuf. Technol.*, vol. 105, no. 9, pp. 3643–3659, Dec. 2019.
[192] M. Monajjemi, S. Mohaimenianpour, and R. Vaughan, "UAV, come to me: End-to-end, multi-scale situated HRI with an uninstrumented human and a distant UAV," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 4410–4417.
[193] Y. Kuno, K. Sadazuka, M. Kawashima, K. Yamazaki, A. Yamazaki, and H. Kuzuoka, "Museum guide robot based on sociological interaction analysis," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Apr. 2007, pp. 1191–1194.
[194] J. Mišeikis, P. Caroni, P. Duchamp, A. Gasser, R. Marko, N. Mišeikiene, F. Zwilling, C. de Castelbajac, L. Eicher, M. Früh, and H. Früh, "Lio—A personal robot assistant for human–robot interaction and care applications," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5339–5346, Oct. 2020.
[195] S. Seneviratne, J. De Zoysa, S. Senarathna, C. Padmasiri, and P. Pallemulla, "Robotic healthcare companion for the elderly and the differently abled with indoor human following and fall detection capability," in *Proc. IEEE 1st Int. Conf. Artif. Intell. Med., Health Care (AIMHC)*, Feb. 2024, pp. 187–193.

[196] W. L. Yeoh, R. Miyata, O. Fukuda, N. Yamaguchi, and H. Okumura, ''Development of a person-following robotic assist Walker with compliant-control arbitrated role-switching,'' *Artif. Life Robot.*, vol. 28, no. 3, pp. 553–559, Mar. 2023.

[197] C. Piezzo and K. Suzuki, ''Design of an accompanying humanoid as a walking trainer for the elderly,'' in *Proc. 25th IEEE Int. Symp. Robot Human Interact. Commun. (RO-MAN)*, Aug. 2016, pp. 467–472.

[198] E. Prassler, D. Bank, and B. Kluge, ''Key technologies in robot assistants: Motion coordination between a human and a mobile robot,'' *Trans. Control, Autom. Syst. Eng.*, vol. 4, no. 1, pp. 56–61, Mar. 2002.

[199] Y. Chen, Z. Xu, Z. Jian, G. Tang, L. Yang, A. Xiao, X. Wang, and B. Liang, ''Quadruped guidance robot for the visually impaired: A comfort-based approach,'' in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2023, pp. 12078–12084.

[200] A. Xiao, W. Tong, L. Yang, J. Zeng, Z. Li, and K. Sreenath, ''Robotic guide dog: Leading a human with leash-guided hybrid physical interaction,'' in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 11470–11476.

[201] S. Cai, A. Ram, Z. Gou, M. A. W. Shaikh, Y.-A. Chen, Y. Wan, K. Hara, S. Zhao, and D. Hsu, ''Navigating real-world challenges: A quadruped robot guiding system for visually impaired people in diverse environments,'' in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2024, pp. 1–18.

[202] G. L. Holmes, K. M. Bonnett, A. Costa, D. Burns, and Y. S. Song, ''Guiding a human follower with interaction forces: Implications on physical human–robot interaction,'' in *Proc. 9th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatronics (BioRob)*, Aug. 2022, pp. 1–6.

[203] N. Sakamoto and M. Okugawa, ''Human following control of porter robot with velocity vectors,'' in *Proc. Dyn. Syst. Control Conf.*, Oct. 2012, pp. 817–821.

[204] S. Thrun, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, ''MINERVA: A second-generation museum tour-guide robot,'' in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 3, Oct. 1999, pp. 1999–2005.

[205] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura, ''The intelligent ASIMO: System overview and integration,'' in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, vol. 3, Oct. 2002, pp. 2478–2483.

[206] A. Al-Wazzan, R. Al-Farhan, F. Al-Ali, and M. El-Abd, ''Tour-guide robot,'' in *Proc. Int. Conf. Ind. Informat. Comput. Syst. (CIICS)*, Mar. 2016, pp. 1–5.

[207] K. Muñoz Peña and B. Bacca Cortes, ''GUI3DXBot: An interactive software tool for a tour-guide mobile robot,'' *Ciencia e Ingeniería Neogranadina*, vol. 30, no. 1, pp. 59–74, Nov. 2019.

[208] J. Leoste, K. Marmor, T. Hollstein, H. Hinkelmann, and L. B. Leoste, ''Enhancing university visitor satisfaction: A human–robot interaction study on the design and perception of a guiding robot assistant,'' in *Proc. Int. Conf. Robot. Educ. (RiE)*. Cham, Switzerland: Springer, Jan. 2024, pp. 223–234.

[209] Y. Chen, F. Wu, W. Shuai, and X. Chen, ''Robots serve humans in public places-kejia robot as a shopping assistant,'' *Int. J. Adv. Robotic Syst.*, vol. 14, no. 3, 2017, Art. no. 1729881417703569.

[210] C. Mavrogiannis, F. Baldini, A. Wang, D. Zhao, P. Trautman,A. Steinfeld, and J. Oh, ''Core challenges of social robot navigation: A survey,'' *ACM Trans. Hum.-Robot Interact.*, vol. 12, no. 3, pp. 1–39, Sep. 2023.

[211] N. Hirose, D. Shah, A. Sridhar, and S. Levine, ''SACSoN: Scalable autonomous control for social navigation,'' *IEEE Robot. Autom. Lett.*, vol. 9, no. 1, pp. 49–56, Jan. 2024.

[212] N. Pérez-Higueras, R. Otero, F. Caballero, and L. Merino, ''HuNavSim: A ROS 2 human navigation simulator for benchmarking human-aware robot navigation,'' *IEEE Robot. Autom. Lett.*, vol. 8, no. 11, pp. 7130–7137, Nov. 2023.

[213] M. Martini, N. Pérez-Higueras, A. Ostuni, M. Chiaberge, F. Caballero, and L. Merino, ''Adaptive social force window planner with reinforcement learning,'' 2024, *arXiv:2404.13678*.

[214] A. Biswas, A. Wang, G. Silvera, A. Steinfeld, and H. Admoni, ''SocNavBench: A grounded simulation testing framework for evaluating social navigation,'' *ACM Trans. Hum.-Robot Interact.*, vol. 11, no. 3, pp. 1–24, Sep. 2022.

[215] M. Fahad, Z. Chen, and Y. Guo, ''Learning how pedestrians navigate: A deep inverse reinforcement learning approach,'' in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 819–826.

[216] R. Firoozi, J. Tucker, S. Tian, A. Majumdar, J. Sun, W. Liu, Y. Zhu, S. Song, A. Kapoor, K. Hausman, B. Ichter, D. Driess, J. Wu, C. Lu, and M. Schwager, ''Foundation models in robotics: Applications, challenges, and the future,'' *Int. J. Robot. Res.*, vol. 2024, Sep. 2024, Art. no. 02783649241281508.

[217] A. Brohan et al., ''RT-1: Robotics transformer for real-world control at scale,'' 2022, *arXiv:2212.06817*.

[218] A. Goyal, J. Xu, Y. Guo, V. Blukis, Y.-W. Chao, and D. Fox, ''RVT: Robotic view transformer for 3D object manipulation,'' in *Proc. Conf. Robot Learn.*, Jan. 2023, pp. 694–710.

[219] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, ''Improving language understanding by generative pre-training,'' Tech. Rep., 2018.

[220] C. Zhang, J. Chen, J. Li, Y. Peng, and Z. Mao, ''Large language models for human–robot interaction: A review,'' *Biomimetic Intell. Robot.*, vol. 3, no. 4, Oct. 2023, Art. no. 100131.

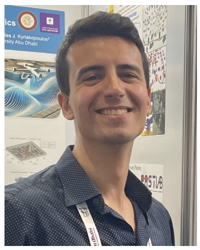
[221] J. Wang, Z. Wu, Y. Li, H. Jiang, P. Shu, E. Shi, H. Hu, C. Ma, Y. Liu, X. Wang, Y. Yao, X. Liu, H. Zhao, Z. Liu, H. Dai, L. Zhao, B. Ge, X. Li, T. Liu, and S. Zhang, ''Large language models for robotics: Opportunities, challenges, and perspectives,'' 2024, *arXiv:2401.04334*.

[222] A. Sorrentino, B. Ferreira, P. Menezes, J. Batista, J. Dias, L. Fiorini, P. Benvenuti, and F. Cavallo, '''Should I lead?' feasibility study of user perception on following-robot for gait assessment,'' in *Proc. 21st Int. Conf. Adv. Robot. (ICAR)*, 2023, pp. 564–569.

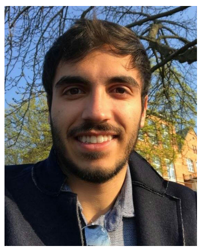
[223] E. Repiso, A. Garrell, and A. Sanfeliu, ''People's adaptive side-by-side model evolved to accompany groups of people by social robots,'' *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2387–2394, Apr. 2020.

[224] M. Li, M. Kwon, and D. Sadigh, ''Influencing leading and following in human–robot teams,'' *Auto. Robots*, vol. 45, no. 7, pp. 959–978, Oct. 2021.

[225] E. Kaufmann, L. Bauersfeld, A. Loquercio, M. Müller, V. Koltun, and D. Scaramuzza, ''Champion-level drone racing using deep reinforcement learning,'' *Nature*, vol. 620, no. 7976, pp. 982–987, Aug. 2023.



ANDREA EIRALE (Member, IEEE) was born in Alba, Cuneo, Italy, in 1996. He received the master's degree in mechatronic engineering from Politecnico di Torino, in 2020, where he is currently pursuing the Ph.D. degree in electrical, electronics, and communication engineering with the PIC4SeR Interdepartmental Centre. His master's thesis was on ''Indoor SLAM and Room Classification with Deep Learning at the Edge.'' His research interests include navigation systems for social service robotics, through traditional methods as well as machine learning approaches.



MAURO MARTINI (Member, IEEE) received the master's degree (cum laude) in mechatronic engineering and the Ph.D. degree in electrical, electronics and communications engineering from the Politecnico di Torino, in 2020 and 2024, respectively, with the thesis ''Machine Learning for Perception and Autonomous Navigation of Service Mobile Robots.'' He is currently carrying out his research activity in collaboration with the Interdepartmental Centre for Service Robotics (PIC4SeR), investigating machine learning solutions for autonomous navigation in-service mobile robotics, with a particular focus on precision agriculture and social navigation.

MARCELLO CHIABERGE (Member, IEEE) is currently an Associate Professor with the Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy, where he is also the Co-Director of the Mechatronics Laboratory and the Director and the Principal Investigator of the Interdepartmental Center for Service Robotics (PIC4SeR). He has authored more than 100 articles accepted in international conferences and journals. He is the co-author of nine international patents. His research interests include the hardware implementation of neural networks and fuzzy systems and the design and implementation of reconfigurable real-time computing architectures.