2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

A Hybrid Human Tracking System using UWB Sensors and Monocular Visual Data Fusion for Human Following Robots

Dingzhi Zhang, Lukas Birner, Felix Pancheri, Christoph Rehekampff, Darius Burschka and Tim C. Lueth, *Senior Member, IEEE*

***Abstract*— The ability to follow people can benefit the human-robot interaction of mobile robots. This work proposes a hybrid human tracking system for human following robots, integrating sensor fusion of Ultra-Wideband (UWB) and monocular visual positioning to enhance tracking accuracy and precision. At the same time, UWB and the visual positioning system can operate independently, thereby creating a redundancy in the system. Based on our previous study of UWB-positioning, this article elaborates on a visual positioning system that employs human detection using a pre-trained Convolutional Neural Network (CNN), coupled with data fusion process based on experimental assessments. The hybrid human tracking system achieves a 2D Euclidean accuracy RMS of 7.4 cm, demonstrating sufficient accuracy for human following and improving the following performance in real-world experiments compared to our previous study.**

I. Introduction

The significance of robotics for society is growing due to demographic changes and resulting labor shortages. User-friendly and collaborative robots can provide essential support in healthcare, help to maintain productivity in industries and assist people in everyday activities. Autonomous robots that follow humans have a wide range of collaborative applications in mobile robotics, including providing transportation assistance in domestic and industrial fields and performing filming tasks in the entertainment industry [1]. In addition, based on the principle of Programming by Demonstration (PbD), human following can also be used as an interface to teach the robot a desired path [2, 3].

The human tracking system plays an important role in human following robots. Based on our implementation of Ultra-Wideband (UWB) based human following in [3], we propose a novel hybrid human tracking system that incorporates a fusion approach using UWB and visual data. In the following, we present the state of the art to distinguish our system from existing approaches.

II. Related Works

Mobile robots with human following capabilities have been widely presented by researchers as early as 1998 [4]. Since then, various approaches have been pursued and different sensors have been used in multiple combinations. This work focuses on human tracking and the associated position estimation relative to the robot. The following review is organized according to the sensor technologies used. First,

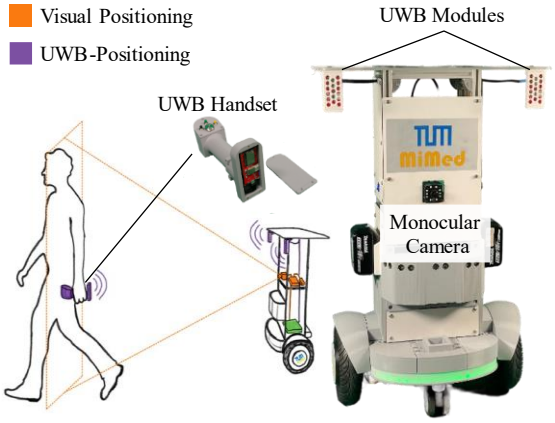

Figure 1. Schematic representation of the UWB-based positioning system consisting of two UWB modules on the robot (anchors) and a UWB module on handset (tag).

non-vision-based techniques are addressed. Then, approaches using only one type of vision sensor are presented, followed by systems using multi-sensor fusion.

Laser range finders (LRF) are frequently utilized in non-vision-based tracking systems. Common approaches include laser data segmentation to find human torsos [5] or human legs [6, 7]. Topp and Christensen [8] allowed the detection of single legs, two legs, and person-wide blobs as human-like shapes. The distance information provided by LRF data are further used to estimate the relative position of the human with respect to the robot. However, these systems may fail when leg features are distorted by long clothes or on occlusions of the target [9].

An alternative method involves the application of radio frequency (RF)-based technologies to acquire position estimates of the human. This method requires the use of a receiver-transmitter device held by the user, thus ensuring a clear distinction of the tracked person. Several wireless technologies have been implemented for human following, including Bluetooth [10], Ultra-Wideband (UWB) technology [3, 11], and Wireless Fidelity (WiFi) [12].

Vision-based systems comprise the majority of human detection and tracking methods. Early approaches to visual human detection include color segmentation and feature detection, which have been employed with various vision sensors, including stereo and RGB-D cameras [4, 13, 14].

This work was supported by School of Engineering and Design, Technical University of Munich, Munich, Germany.

Dingzhi Zhang, Lukas Birner, Felix Pancheri, Christoph Rehekampff and Tim C. Lueth are with the Institute of Micro Technology and Medical Device Technology (MiMed), Technical University of Munich, Munich, Germany.

Darius Burschka is with the Machine Vision and Perception Group, Department of Computer Science, Technical University of Munich, Munich Germany.

Contact: Dingzhi.Zhang@tum.de

More recently, machine learning algorithms have been utilized for human detection and tracking in vision-based applications [15, 16]. However, vision-based human tracking systems may experience performance degradation in challenging situations, such as cluttered backgrounds or difficult lighting conditions [9].

To address the limitations of vision-based systems, approaches have been made to combine vision sensors with both LRF and RF-based sensors. Alvarez-Santos *et al.* [17] used a monocular camera and an LRF. The camera detected a human's torso and its angle relative to the robot by utilizing a gradient-based detection algorithm in combination with a feature model of the target's torso for tracking. The LRF performed leg detection and produced distance estimates. Sarmento *et al.* [18] combined monocular vision and UWB transceivers for human position estimates. However, the camera was used to reduce the angular error determined by UWB sensors but cannot perform full 2D tracking.

III. Proposed Approach

This work introduces a hybrid human tracking system using UWB sensors and a monocular camera for human following robots. The primary benefits of the proposed system include: 1) The use of both sensor types eliminates the need for direct visual contact and ensures unambiguous identification of the person tracked. 2) In contrast to [17, 18], each sensor is capable of operating independently as a standalone positioning system, determining the relative distance and angular difference between the human operator and a robot. 3) In addition to positioning, the UWB handset allows remote control of the robot during operation.

A. Hardware Selection and Design

The proposed hybrid approach for human tracking combines UWB-based and visual positioning. Fig. 1 illustrates the UWB positioning system, consisting of three UWB modules. Two of these modules are mounted on the robot and are connected to the microcontroller via I2C (see Fig. 2). The third module is housed in a handset. Using the Two-Way Ranging (TWR) described in [3], the position of the handset is determined relative to the robot. The handset also contains buttons that allow the operator to remotely control the robot. In a following scenario, when the operator's back faces the robot, the handset is designed to orient its UWB module towards the UWB modules mounted on the robot, ensuring optimal positioning performance.

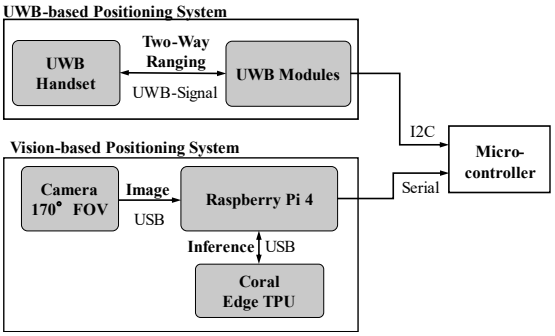

Figure 2. Architechture of the hybrid human tracking system consisting of UWB-based and vison-based positioning.

The UWB positioning system is complemented by human recognition using a wide-angle monocular camera mounted on the robot. The maximum field of view (FOV) of 170 degrees of the USB camera enables the human operator to be visible at the desired following distance between 1 and 2.2 m. This is considered a comfortable distance range [19]. As shown in Fig. 2, human detection is based on inference from a pre-trained CNN using the Google Coral USB Edge TPU. The Raspberry Pi is connected to the microcontroller via UART. This allows position estimates obtained by the visual positioning system to be sent to the microcontroller, where the sensor fusion is performed.

B. Visual Positioning System

The visual positioning system estimates the relative position between the robot and the human operator using object detection. The detection and tracking pipeline, illustrated in Fig. 3, runs on a Raspberry Pi 4 at an average speed of 17 frames per second.

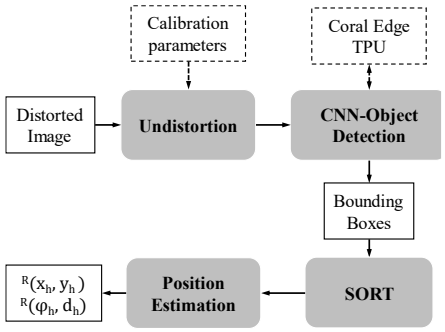

Figure 3. Human detection and tracking pipeline of the visual positioning system.

The process starts by capturing a new frame from the camera stream. The image is then undistorted using pre-determined calibration parameters to correct distortion caused by the camera lens. Afterward, the frame is resized to match the input size of the selected CNN model, which is 320 x 320 pixels across three color channels. A MobileDet model of Single Shot MultiBox Detector Lite (SSDLite) is used, which has been pre-trained on the Common Objects in Context (COCO) dataset [20] and compiled for execution on the Coral Edge TPU. The inference returns bounding boxes with associated class labels and detection scores for the detected objects. Only human detections are filtered. The tracking algorithm SORT [21] assigns a unique tracking ID to the bounding boxes of each human detection and updates the IDs accordingly. These tracks are then used to obtain a position estimate of the person with the smallest associated track ID. Therefore, during the initialization of the visual tracking system, it is necessary that the operator is the first detectable person in the frame. After the human is detected and tracked, the position of the human relative to the robot is estimated using the coordinates of the bounding box. This position is expressed both in polar coordinates $^{R}(\varphi_h, d_h)$ or in the Cartesian coordinates $^{R}(x_h, y_h)$ in the robot's frame **R**. An example of an undistorted image frame at the end of the detection and tracking process is shown in Fig. 4 a). The polar coordinates next to the green bounding box show the estimated position of the person's heel, which is marked with a red dot.

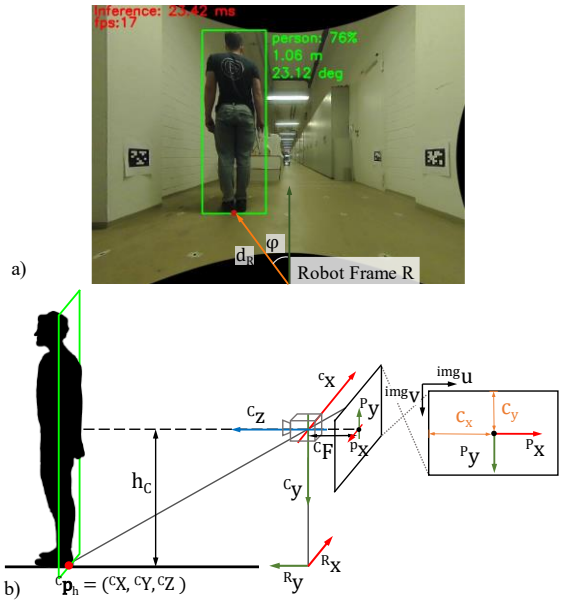

Figure 4. Visual positiong based on human detection. a) Example frame from visual detection and tracking with operator and associated tracking results. b) Geometric relations for visual positioning

The human operator's position estimation is based on the relationship between the person's heel $\mathbf{p}_h$ and its projection in the image frame, which is the horizontal center of the lower edge of the bounding box (red dot in Fig. 4). The coordinates of point $\mathbf{p}_h$ in the camera frame **C** are $^C\mathbf{p}_h(^cX,^cY,^cZ)$. Its projection in the image plane coordinate **P** and in the image pixel coordinate **img** are $^P\mathbf{p}_h(x,y)$ and $^{img}\mathbf{p}_h(u,v)$ respectively. Since the image frame has already been undistorted as part of the detection and tracking pipeline described above, the pinhole camera model can be assumed. Therefore, projection equations apply between **C** and **P**:

$$^P x = \frac{^C F}{^C Z}\, ^C X \tag{1}$$

$$^P y = \frac{^C F}{^C Z}\, ^C Y \tag{2}$$

^{C}F is the camera's focal length in length units of **C**. The task of the visual positioning system is to determine the position of $^C\mathbf{p}_h$ based on the image data after the detection and tracking pipeline, which is available in pixel coordinate of **img**. Thus, the transformation of the image frame **img** to the camera frame **C** is required, which can be expressed as:

$$^C\boldsymbol{T}_{img} = {}^C\boldsymbol{T}_P \cdot {}^P\boldsymbol{T}_{img} \tag{3}$$

$^C\mathbf{T}_P$ is the 3D to 2D projection from the camera frame **C** into the image plane coordinate **P** given by (1-2). $^P\mathbf{T}_{img}$ describes the transformation between **P** and **img**, which contains only a translation given by the principal point position in pixels (C_x, C_y) in the **img** coordinate. Thus, the relationship between $^C\mathbf{p}_h(X,Y,Z)$ and $^{img}\mathbf{p}_h(u,v)$ is:

$$^{img}u = {}^{img}f_x \frac{^C X}{^C Z} + c_x \tag{4}$$

$$^{img}v = {}^{img}f_y \frac{^C Y}{^C Z} + c_y \tag{5}$$

with $^{img}f_x$ and $^{img}f_y$ referring to the focal length given in the pixel size in x and y direction. Note that the camera is mounted on the robot such that the axes of the robot frame **R** are congruent with the axes of **C**.

In (4) and (5), $^{img}(u,v)$ are available from the bounding box obtained by the detection and tracking pipeline. The parameters (f_x, f_y, c_x, c_y) along with the distortion parameters are determined by the camera calibration using an appropriate camera model. The problem of position estimation is to compute the world coordinate of point P_h given its pixel values. However, with two equations and three unknowns $^C(X,Y,Z)$, it cannot be solved. Thus, it is assumed that the operator touches the ground while walking, and P_h is a sufficiently accurate estimate of the heel's position in the real world. Therefore, the Y-coordinate of $^C\mathbf{p}_h$ is known to be the camera's height above the ground h_c as indicated in Fig.4 b). From these considerations, the following two equations are formulated and give the heels Z- and X-coordinates in **C** in dependence of h_c, the focal lengths (f_x, f_y), the coordinates of the principal point (c_x, c_y), and the image coordinates $^{img}(u,v)$:

$$^C Z_h = f_y \frac{h_c}{(^{img}v_h - c_y)} \tag{6}$$

$$^C X_h = \frac{^C Z_h}{f_x}(^{img}u_h - c_x) \tag{7}$$

The positioning results are filtered using moving median filters for X_h and Z_h, each with a window size of 10 to account for measurement noise and temporary track losses. Eventually, the coordinates of $\mathbf{p}_h$ are converted to polar coordinates and Cartesian coordinates in the robot's frame **R**.

C. Sensor Fusion of UWB and Visual Data

The fusion of two independent position estimates from the UWB positioning and the visual positioning system requires a scheme for effectively merging both results to enhance the performance of the combined system. This section presents the motivation and derivation of the implemented fusion algorithm.

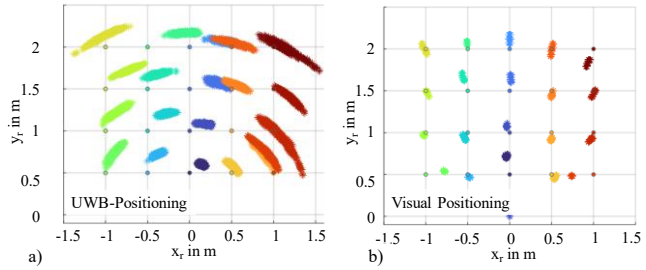

Figure 5. One of the five test measurement sequences of the a) UWB-based positioning system and b) visual positioning system. The circles at the intersections of the grid are reference points. Identical colors indicate the correspondence between the reference and measurement points.

Test measurements were conducted separately with both positioning systems to evaluate their properties and determine an efficient scheme for sensor fusion. The robot was positioned at a fixed location (0,0), and a grid of markers was placed on the ground in a 2 × 2 m rectangle at 0.5 m intervals in front of the robot. The operator held the handset in his right hand close to his body and was standing on each marker for ten seconds while both positioning systems recorded their

respective measurements. This process was repeated five times. One of these measurement results is shown in Fig. 5.

The following qualitative conclusions can be made based on a visual inspection of the results.

- The UWB positioning system exhibits high noise in angle measurements.
- The visual positioning system measures angle with high precision compared to the UWB positioning system, but still encounters fluctuations in distance estimation.

TABLE I. AVERAGE OF MEAN POLAR ERRORS OVER FIVE TEST MEASUREMENT SEQUENCES

UWB		**Vision**	
$\mu_{\varphi,\,u}$ (rad)	$\mu_{d,\,u}$ (m)	$\mu_{\varphi,\,v}$ (rad)	$\mu_{d,\,v}$ (m)
-0.138	0.055	$-2.6\text{x}10^{-3}$	$0.1\text{x}10^{-3}$

The results suggest that a complementary sensor configuration is advantageous when considering polar coordinates (see Fig.4 a)). The quantitative evaluation of the sample measurements also supports the qualitative results mentioned above. The angle and distance deviation, φ and d, which determine the average polar errors, were calculated for both sensors across five measurement sequences (cf. Tab. I). Additionally, Tab. II displays the corresponding medians of the inverse variances for the polar errors. When calculating the variance of the angular deviations over the 20 measurement positions within one test sequence, it is preferable to use the median instead of the mean. This is because at certain measuring positions, the angular deviation of the vision system has a standard deviation σ close to zero, which could result in undefined values for $1/\sigma^2$.

TABLE II. AVERAGE OF MEDIAN VALUES OF INVERSE VARIANCE OF POLAR ERRORS OVER FIVE TEST MEASUREMENT SEQUENCES

UWB		**Vision**	
$1/\sigma^2_{\varphi,u}$	$1/\sigma^2_{d,u}$	$1/\sigma^2_{\varphi,v}$	$1/\sigma^2_{d,v}$
670	7932	62255	3743

The fusion algorithm for the UWB and visual positioning data utilizes the inverse-variance weighting average [22]. It assumes that the polar measurements from the UWB and from the visual positioning system have a normally distributed measurement noise.

The hybrid system conducts the sensor fusion in polar coordinates. Additionally, angle φ and distance d should be evaluated separately. For this, Eq. (8-9) are utilized: once to obtain a fused estimate for angle φ and a second time to evaluate distance d. Initially, each measurement is corrected for the sensor bias represented by the mean of their corresponding measurement noises μ_u and μ_v. Subsequently, the corrected measurements are weighted according to the inverse variance of the corresponding measurement noise $1/\sigma_u^2$ and $1/\sigma_v^2$.

$$\hat{\varphi} = \frac{\frac{1}{{\sigma_{\varphi,u}}^2}(\varphi_u - \mu_{\varphi,u}) + \frac{1}{{\sigma_{\varphi,v}}^2}(\varphi_v - \mu_{\varphi,v})}{\frac{1}{{\sigma_{\varphi,u}}^2} + \frac{1}{{\sigma_{\varphi,v}}^2}} \tag{8}$$

$$\hat{d} = \frac{\frac{1}{{\sigma_{d,u}}^2}(d_u - \mu_{d,u}) + \frac{1}{{\sigma_{d,v}}^2}(d_v - \mu_{d,v})}{\frac{1}{{\sigma_{d,u}}^2} + \frac{1}{{\sigma_{d,v}}^2}} \tag{9}$$

The fusion parameters required are determined from the five test measurement sequences described in the previous section. Tab. III provides a summary of these parameters.

TABLE III. SUMMARY OF PARAMETERS FOR SENSOR FUSION

Biases φ (rad), d (m)				**Inverse variance**			
UWB		***Vision***		***UWB***		***Vision***	
$\mu_{\varphi,\,u}$	$\mu_{d,\,u}$	$\mu_{\varphi,\,v}$	$\mu_{d,\,v}$	$\frac{1}{\sigma^2_{\varphi,u}}$	$\frac{1}{{\sigma^2_{d,u}}^2}$	$\frac{1}{\sigma^2_{\varphi,v}}$	$\frac{1}{\sigma^2_{d,v}}$
0	0.055	0	0	670	7932	62255	3743

The confidence values are derived from the corresponding averaged results presented in Tab. II. Regarding the biases, only one parameter is chosen to be non-zero. That is, $\mu_{d,\,u}$, which accounts for the distance bias of the UWB sensor that was identified during the test sequences. Although there may be biases in the angle measurements of this sensor system, they are set to zero in the fusion process. These are primarily due to the operator holding the UWB tag in his right hand. The fusion model is designed to work independently of the operator's handedness. Since the angle measurement from the UWB has a small contribution to the fused angle estimate, it is neglected. The average biases for the angle and distance of the vision system are less than 0.003 rad and 0.02 cm, respectively, and are therefore also neglected.

IV. EXPERIMENTAL RESULTS

This work presents a hybrid human tracking system using sensor fusion of UWB and visual data for human-following robots. Common indoor environments, such as those found in offices or hospitals, often contain obstructions, sharp turns, and narrow corridors. Therefore, an accurate following behavior of the robot is desirable. The quality of the follow mode is related to the accuracy of the position estimates from the positioning system. In this section, we first evaluate the accuracy of the proposed hybrid positioning system. Then, we assess the quality of human following using our in-house mobile platform *Carrier Bot*.

A. Accuracy of the Hybrid Positioning System

The first experiment aims to quantitatively evaluate the measurement accuracy of the hybrid positioning system. The robot is placed in a fixed position, from which its coordinate frame can be derived, as depicted in Fig. 6. Both anchors of the UWB positioning system are mounted underneath the top tray at a fixed distance of d = 0.48 m. The operator holds the handset in his right hand. The camera for the visual positioning system is mounted at the center of the robot's frame, facing forward. Markers on the floor indicate the measurement positions for the operator during the recording of the samples. The reference positions are shown in Fig. 6 b). Each position is measured 500 times at a sampling frequency of 50 Hz.

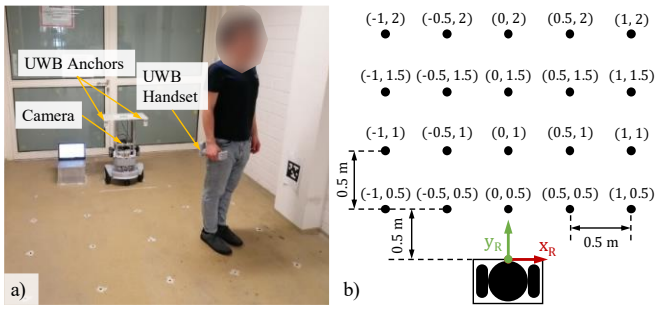

Figure 6. Experiment to determine the positioning accuracy. a) Expriment setup b) Measurement positions

The measurement results of the hybrid system are given in Fig. 7. The reference positions are depicted as colored circles. Measurement recorded at the corresponding makers are shown as point clouds of the same color.

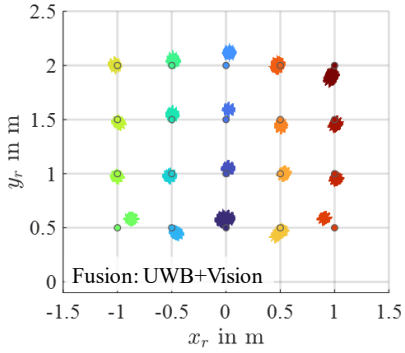

Figure 7. Measurement results of the hybrid positioning system using UWB and visual data fusion.

In addition to the fusion results, measurements only from UWB and vision system are also recorded for comparison. The mean of the Euclidian distance errors $\bar{d}$ using the three different methods are given in Tab. IV. Standard deviation σ of each system is also calculated to measure the dispersion. It shows that the mean Euclidean error and the variance of the combined system is less than that of each individual system. Within the measurement area, the hybrid approach reduces the Euclidean error by 74% and the standard deviation by 73% in contrast to the UWB-based positioning alone. Compared to the UWB positioning system based on Phase-Difference-of-Arrival (PDOA), the proposed hybrid system shows similar distance accuracy and has better stability regarding orientation-dependent fluctuations [11]. Despite using cost-effective hardware with a total cost around 300 €, our proposed system achieved similar accuracy with an average distance error of 6.3 cm compared to the fusion-based method consisting of UWB and vision in [18] with an average distance error of 5.4 cm. The cost of the camera alone used in [18] would be similar to the cost of the entire hybrid system in our implementation.

TABLE IV. AVERAGE OF EUCLIDEAN DISTANCE ERRORS AND STANDRD DEVIATIONS BETWEEN SINGLE AND HYBRID SYSTMEMS

UWB		Vision		Fusion	
μ	σ	μ	σ	μ	σ
24 cm	13.9 cm	10.6 cm	8.5 cm	6.3 cm	3.8 cm

Since the proposed tracking system is to be used for human following robots, the question arises how accurate the estimates need to be. According to [23], the average stride width of human gait is 0.1 m, measured as the distance between the centerlines of the soles. When combined with the average foot breadth of 0.1 m [24, 25], the inherent imprecision of a human path is assumed to be ±0.1 m from the centerline of a hypothetical path. Therefore, a positioning system that can determine human position with an accuracy of at least 0.1 m is considered satisfactory for human-following applications. Using a one-tailed t-test, the distance error is determined to be less than 0.1 m with a significance level of $\alpha = 0.05$ with $n = 1000$ samples. The overall 2D-RMSE of the position estimate is 7.4 cm.

B. *Quality of Human Following*

The following section evaluates the behavior of the robot using the implemented tracking system, with a measurement frequency of 20 Hz. The evaluation is divided into two parts: the robot's ability to maintain a comfortable following distance and its accuracy in following a human path. Fig. 8 displays the setup and results of both experiments.

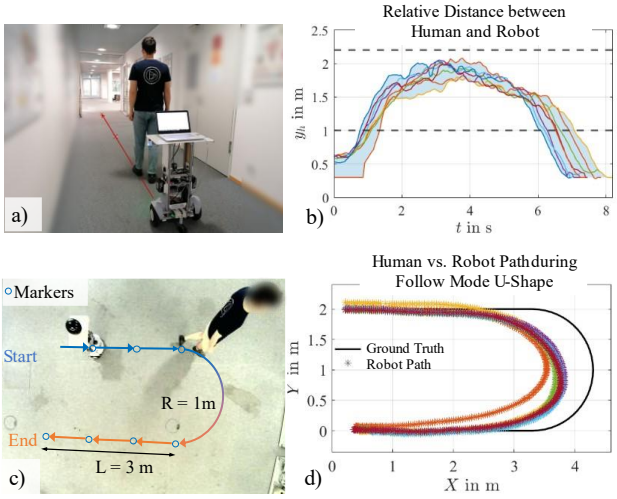

Figure 8. Experiments for the evaluation of the robot's following behavior. a) Experiment setup for measuring following distance. b) Following distance recorded during a linear path. c) Experiment setup for measuring accuracy of path following. d) Comparison of robot path and human path.

In the first experiment, the robot followed the human operator eight times along a 10 m linear path that is indicated by the markers on the ground. The robot's onboard sensors recorded the distance ${}^{R}y_h$ to the human. According to [19], a comfortable following distance can be considered between 1-2 m. These thresholds are represented by the dashed lines in Fig. 8 b). The results reveal that the conducted samples were mostly within these limits, with exceptions at the beginning and the end of the following motion. This corresponds to the implemented behavior, namely the activation of the follow mode at a distance of 0.5 m and the gradual approach to the human operator at the end of the following motion.

To evaluate the quality of following a human path, a U-shaped path with a curvature radius of 1 m was used. Six

measurements using a camera-based tracking system were conducted. Fig. 8 d) shows the operator's path in black, which is assumed to be the reference path, along with the robot's tracked positions during the measurements. It can be observed that the robot repeatedly cut the corner and chose a shorter path, resulting in deviations from the operator's path. This behavior is primarily due to the implemented controller. In contrast to using only UWB sensors for human following as described in [3], deviations in straight lines are barely noticeable. This reflects the higher precision of the hybrid positioning system (see Tab. IV). Euclidean distances of the nearest point between the reference path and the robot path are further calculated. The average distance was 0.21 m, and the average maximum distance over six measurements was 0.6 m.

V. Conclusion and Future work

A hybrid human tracking system based on UWB and visual positioning for human following robots has been realized using cost-effective hardware in this work. Experimental evaluation demonstrates that the system achieves accuracy with 2D-RMSE of less than 7.4 cm, which is sufficient for a human following robot. The fusion-based approach improves the human following performance of our in-house mobile platform *Carrier Bot*. In particular, it reduces the robot's lateral fluctuation due to the enhanced precision of the hybrid positioning system. Future studies may consider implementing situational adaptive follow modes, such as side-follow or back-follow. In addition, improvements to the follow controller can be made by integrating human walking trajectory, obstacle avoidance and considering the rules of social space as proposed in [26]. This could improve human path imitation and make teach-and-repeat path programming more intuitive.

References

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

[2] V. Alvarez-Santos, A. Canedo-Rodriguez, R. Iglesias, X. M. Pardo, C. V. Regueiro, and M. Fernandez-Delgado, "Route learning and reproduction in a tour-guide robot," Robotics and Autonomous Systems, vol. 63, pp. 206–213, 2015, doi: 10.1016/j.robot.2014.07.013.

[3] D. Zhang, L. Birner, V. Zinkernagel, C. Rehekampff, D. Burschka, and T. C. Lueth, "Carrier Bot: A UWB-based Human Following Mobile Platform for Intra-Office Transport with an Intuitive Teach-and-Repeat Programming," in 2023 IEEE International Conference on Robotics and Biomimetics (ROBIO), Koh Samui, Thailand, 2023, pp. 1–6.

[4] C. Schlegel, J. Illmann, H. Jaberg, M. Schuster, and R. Wörz, "Vision Based Person Tracking with a Mobile Robot," in Procedings of the British Machine Vision Conference 1998, Southampton, 1998, 42.1-42.10.

[5] N. A. Olmedo, H. Zhang, and M. Lipsett, "Mobile robot system architecture for people tracking and following applications," in 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014), Bali, Indonesia, 2014, pp. 825–830.

[6] A. Leigh, J. Pineau, N. Olmedo, and H. Zhang, "Person tracking and following with 2D laser scanners," in 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 2015, pp. 726–733.

[7] H. Yao, H. Dai, E. Zhao, P. Liu, and R. Zhao, "Laser-Based Side-by-Side Following for Human-Following Robots," in 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 2021, pp. 2651–2656.

[8] E. A. Topp and H. I. Christensen, "Tracking for following and passing persons," in 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Alta., Canada, 2005, pp. 2321–2327.

[9] K. Koide and J. Miura, "Identification of a specific person using color, height, and gait features for a person following robot," Robotics and Autonomous Systems, vol. 84, pp. 76–87, 2016, doi: 10.1016/j.robot.2016.07.004.

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

[11] 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 2022 IEEE International Conference on Robotics and Biomimetics (ROBIO), Jinghong, China, 2022, pp. 333–338.

[12] V. Geetha, S. Salvi, G. Saini, N. Yadav, and R. P. Singh Tomar, "Follow Me: A Human Following Robot Using Wi-Fi Received Signal Strength Indicator," in Advances in Intelligent Systems and Computing, ICT Systems and Sustainability, M. Tuba, S. Akashe, and A. Joshi, Eds., Singapore: Springer Singapore, 2021, pp. 585–593.

[13] 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, 2017, doi: 10.1109/TSMC.2016.2616343.

[14] T. Yoshimi et al., "Development of a Person Following Robot with Vision Based Target Detection," in 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, 2006, pp. 5286–5291.

[15] B. X. Chen, R. Sahdev, and J. K. Tsotsos, "Integrating Stereo Vision with a CNN Tracker for a Person-Following Robot," in Lecture Notes in Computer Science, Computer Vision Systems, M. Liu, H. Chen, and M. Vincze, Eds., Cham: Springer International Publishing, 2017, pp. 300–313.

[16] K. Koide and J. Miura, "Convolutional Channel Features-Based Person Identification for Person Following Robots," in Advances in Intelligent Systems and Computing, Intelligent Autonomous Systems 15, M. Strand, R. Dillmann, E. Menegatti, and S. Ghidoni, Eds., Cham: Springer International Publishing, 2019, pp. 186–198.

[17] V. Alvarez-Santos, X. M. Pardo, R. Iglesias, A. Canedo-Rodriguez, and C. V. Regueiro, "Feature analysis for human recognition and discrimination: Application to a person-following behaviour in a mobile robot," Robotics and Autonomous Systems, vol. 60, no. 8, pp. 1021–1036, 2012, doi: 10.1016/j.robot.2012.05.014.

[18] J. Sarmento, F. Neves dos Santos, A. Silva Aguiar, V. Filipe, and A. Valente, "Fusion of Time-of-Flight Based Sensors with Monocular Cameras for a Robotic Person Follower," J Intell Robot Syst, vol. 110, no. 1, 2024, doi: 10.1007/s10846-023-02037-4.

[19] F. W. Siebert, J. Klein, M. Rötting, and E. Roesler, "The Influence of Distance and Lateral Offset of Follow Me Robots on User Perception," Frontiers in robotics and AI, vol. 7, p. 74, 2020, doi: 10.3389/frobt.2020.00074.

[20] Y. Xiong et al., "MobileDets: Searching for Object Detection Architectures for Mobile Accelerators," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021, pp. 3824–3833.

[21] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 2016, pp. 3464–3468.

[22] J. Hartung, G. Knapp, and B. K. Sinha, Statistical meta-analysis with applications. Hoboken, NJ: Wiley, 2008. [Online]. Available: http://www.loc.gov/catdir/enhancements/fy0811/2008009435-d.html

[23] L. Proksch, M.-M. Véghelyi, and A. Sturma, "Veränderungen im Gangbild bei Personen mit unilateraler transtibialer Amputation – Ein systematisches Literaturreview," physioscience, vol. 16, no. 02, pp. 72–84, 2020, doi: 10.1055/a-1114-1975.

[24] K. Baba, "Foot measurement for shoe construction with reference to the relationship between foot length, foot breadth, and ball girth," Journal of human ergology, vol. 3, no. 2, pp. 149–156, 1974.

[25] M. R. Hawes and D. Sovak, "Quantitative morphology of the human foot in a North American population," Ergonomics, vol. 37, no. 7, pp. 1213–1226, 1994, doi: 10.1080/00140139408964899.

[26] J. Peng, Z. Liao, Z. Su, H. Yao, Y. Zeng, and H. Dai, "A Dual Closed-Loop Control Strategy for Human-Following Robots Respecting Social Space," in 2024 IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 2024, pp. 11252–11258.