A Multi-sensor Combined Tracking Method for Following Robots

Hao Liu, Gang Yu(✉), and Han Hu

Department of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen, Shenzhen 518055, China
gangyu@hit.edu.cn

Abstract. At present, the research on tracking methods is mainly based on visual tracking algorithm, which has reduced the accuracy at night or under the condition of insufficient light intensity. Therefore, this paper starts from the direction of multi-sensor combined tracking. Firstly, in order to verify the feasibility and performance of the multi-sensor combined tracking method proposed in this paper, a set of tracking robot system is designed. Secondly, aiming at the problem that the visual tracking method fails to track in scenes such as complete occlusion and insufficient illumination, the non-line-of-sight perception of the following target is realized based on the fusion of ultra-wide band (UWB) and inertial measurement unit (IMU) sensors. Besides, based on coordinate transformation and decision tree algorithm, this paper makes decisions on UWB and visual tracking targets to achieve combined tracking.

Keyword: Following robot · Target tracking · Multi-sensor combined tracking

1 Introduction

The Intelligent System Laboratory of the Central Research Institute of Toyota in Japan has developed a personal following robot to assist in handling and loading [1]. The following robot is mainly equipped with panoramic camera, Lidar and inertial measurement sensor for sensing, and the robot follows the target according to the robot's kinematic model. Among them, the panoramic camera is used for target tracking, Lidar is used for obstacle avoidance, and inertial sensors measure the acceleration and angular acceleration of the robot to control the robot to keep its balance.

The school of robotics engineering at Inha University in Korea has developed a following robot [2]. The following robot mainly integrates monocular camera and Lidar to track human targets. In the aspect of visual tracking, the particle filter method is used to track the morphological features extracted from the image. At the same time, the laser ranging sensor is used to measure the distance and angle of the target, and then the data of Lidar and visual tracking are fused to realize the reliable tracking of the target.

Han yang University in South Korea has developed a following robot for marathon athletes [3]. The robot mainly obtains the point cloud data of the surrounding environment through laser sensors. According to the support vector data description, the point

H. Liu et al. (Eds.): ICIRA 2022, LNAI 13456, pp. 722–734, 2022.
https://doi.org/10.1007/978-3-031-13822-5_65

cloud data is mapped to high-dimensional space for classification, and the target area is distinguished, the target position is tracked. At the same time, Kalman filter is used to estimate the state of the tracking human body and the optimal position of the tracking target, to realize the motion control of the robot.

The commercial version developed by Intel follows the Segway robot. An intelligent upgrade is made on the platform of the balance car, which senses the surrounding environment through the RGB-D camera, realizes gesture recognition, obstacle avoidance and following based on the visual algorithm, and also has the functions of speech recognition, mobile photography and home monitoring.

A humanoid robot developed by Shenyang Institute of automation, Chinese Academy of Sciences [4], which realizes target tracking based on three degrees of freedom redundant vision. It is equipped with a binocular camera composed of a laser based TOF camera and two CCD cameras. Its processing logic is very similar to human eyes, which is to find and track the target in a relatively large range. After finding the target, carefully observe the target and track it. Firstly, it looks for the target to be tracked through the time-of-flight camera, roughly locates the target, and then accurately locates it through the binocular camera. The time-of-flight camera does not have a high resolution, which can reduce the computation during the coarse localization phase. However, the binocular camera has a higher resolution, which allows precise measurement and localization of the target.

2 Following System Design

The method of human target tracking based on multi-sensor proposed in this paper is mainly used to solve the shortcomings of visual tracking. For example, in completely obscured and poorly illuminated scenes, the vision tracking algorithm is unable to re-identify and track the target when the human target leaves the camera's field of view. Therefore, this paper introduces Ultra Wide Band (UWB) and Inertial Measurement Unit (IMU) sensors, which mainly address the problem of how to provide reliable coordinate information for tracking targets in the presence of occlusion.

Based on the analysis of the above application scenarios and adopted technologies, the overall scheme design of the following robot in this paper is mainly divided into four parts, including core processing layer, hardware layer, control system layer and power module as shown in Fig. 1.

(1) The perception layer is the bottom part of the whole system, and it is also an important device for following robot to realize the perception of the surrounding environment. The 1080p camera sensor is installed on the 2-DOF camera PTZ (Pan/Tilt/Zoom) to provide the system with video stream in the following process. Two nine axis gyroscopes are respectively installed on the camera pan tilt and the chassis of the following robot to provide the system with the relative angle and pitch angle of the camera. The tags of IMU and UWB are fixed together with Bluetooth module to provide the acceleration information and angle information of human target for the system. The motor driver is responsible for controlling the motor following the robot motion.

(2) The core processing layer is built based on ROS system, which is mainly responsible for visual target detection and tracking and sensor combination tracking. In the initialization process, the target detection algorithm detects the human target closest to the robot in the picture as the target tracked in the subsequent process, and releases the detected target. The target tracking node subscribes to the target location data of the target detection node as the initialization target box of target tracking. The fusion node parses the data according to the communication protocol and obtains the UWB tracking data and the attitude data of the nine-axis gyroscope sensor. At the same time, it achieves the matching between the UWB tracking target and the visual tracking target according to the conversion relationship between the UWB coordinate system and the camera coordinate system, and achieves the effect of multi-sensor tracking fusion.

(3) The control system layer mainly collects the sensing data with the core processing layer through UART serial port and transmits it to the core processing layer for processing. There are two processing units in the control layer. The first processing unit STM32F103 is responsible for processing the measurement data of each UWB base station and label. Through the distance data between the three base stations and the label, the position information of the label relative to the robot can be obtained through the solution algorithm. Another processing unit STM32F407 is mainly responsible for the PTZ attitude control of the 2-DOF camera. It controls the motor speed of the robot chassis through the motor driver and receives the data from the core processing layer. At the same time, it receives the position information calculated from STM32F103 and sends the relevant data to the core processing unit through the serial port.

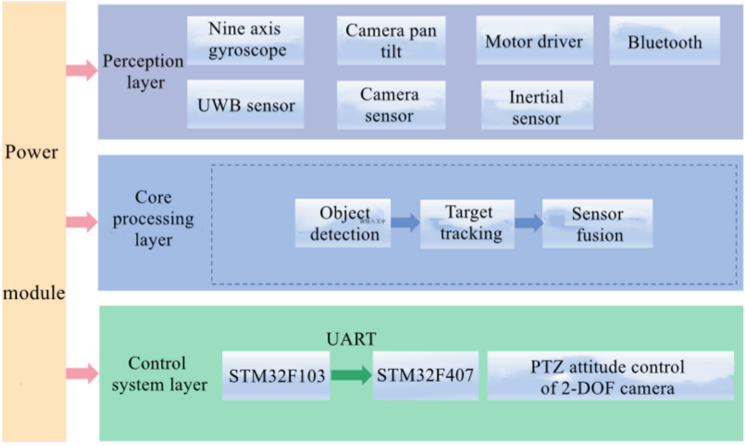

Fig. 1. Overall scheme design of following robot

The hardware layout of the following robot is shown in Fig. 2.

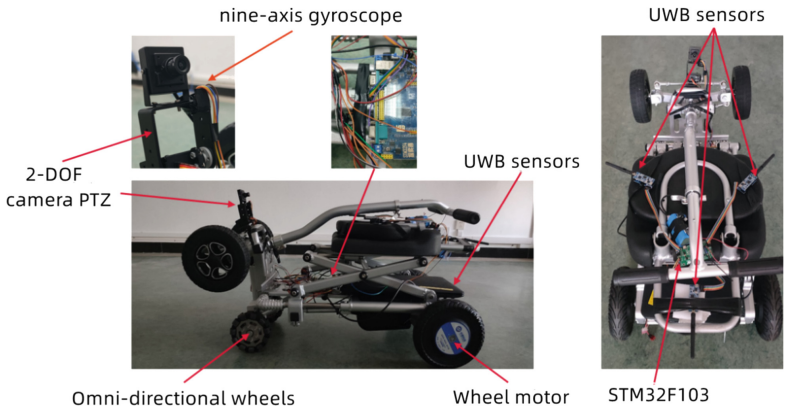

Fig. 2. The hardware layout of the following robot

3 Multi-sensor Combined Tracking

3.1 Fusion Model of UWB and IMU Based on Adaptive Kalman Filter

According to Kalman filter, for general linear system, the relationship between state equation and observation equation is as follows:

$$X_k = \Phi_{k,k-1} X_{k-1} + B_{k-1} U_{k-1} + \Gamma_{k-1} W_{k-1} \tag{1}$$

$$Z_k = H_k X_k + V_k \tag{2}$$

where X_k represents state vector at time k, B_{k-1} represents the influence of $k-1$ time input on the system, U_{k-1} represents $k-1$ time input, W_{k-1} represents dynamic noise of random system, H_k represents k-time measurement matrix, V_k represents measurement noise sequence at time k, Γ_{k-1} represents System noise matrix.

In this paper, the distance measured by three base stations $[d_1, d_2, d_3]^T$ is directly used as the observation variable, that is, the observation equation is nonlinear and needs to be processed by extended Kalman filter. According to the human target model in this paper, it can be obtained by Kalman filter. We can get the state vector describing human motion and the measurement of human target view obtained by using UWB solution model as follows:

$$\begin{pmatrix} x_k \\ y_k \\ \dot{x}_k \\ \dot{y}_k \end{pmatrix} = \begin{pmatrix} I_2 & \Delta T I_2 \\ 0 & I_2 \end{pmatrix} \begin{pmatrix} x_{k-1} \\ y_{k-1} \\ \dot{x}_{k-1} \\ \dot{y}_{k-1} \end{pmatrix} + \begin{pmatrix} \frac{1}{2}\Delta T^2 I_2 \\ \Delta T I_2 \end{pmatrix} \begin{pmatrix} \ddot{x}_{k-1} \\ \ddot{y}_{k-1} \end{pmatrix} + W_{k-1} \tag{3}$$

$$\begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = \begin{pmatrix} \sqrt{(x_k - x_1)^2 + (y_k - y_1)^2} \\ \sqrt{(x_k - x_2)^2 + (y_k - y_2)^2} \\ \sqrt{(x_k - x_3)^2 + (y_k - y_3)^2} \end{pmatrix} + V_k \tag{4}$$

where, ΔT represents the sampling time interval, I_2 represents the 2×2 unit matrix, the positions of the three UWB base stations are (x_1, y_1),(x_2, y_2) and (x_3, y_3), respectively, and (x_k, y_k) represents the position of the tracked human body.

$$h(X_k) = \begin{pmatrix} \sqrt{(x_k - x_1)^2 + (y_k - y_1)^2} \\ \sqrt{(x_k - x_2)^2 + (y_k - y_2)^2} \\ \sqrt{(x_k - x_3)^2 + (y_k - y_3)^2} \end{pmatrix} \tag{5}$$

Because $h(X_k)$ is a nonlinear function, there is no constant matrix H_k, which makes both sides of the equation hold. According to the extended Kalman filter, the nonlinear function is expanded by Taylor formula, and the approximate linearized equation is obtained by ignoring the higher-order terms of more than quadratic. Then you can get:

$$H_k = \frac{\partial h(X_k)}{\partial X_k} = \begin{pmatrix} \frac{\partial p_1}{\partial x_k} & \frac{\partial p_1}{\partial y_k} & \frac{\partial p_1}{\partial \dot{x}_k} & \frac{\partial p_1}{\partial \dot{y}_k} \\ \frac{\partial p_2}{\partial x_k} & \frac{\partial p_2}{\partial y_k} & \frac{\partial p_2}{\partial \dot{x}_k} & \frac{\partial p_2}{\partial \dot{y}_k} \\ \frac{\partial p_3}{\partial x_k} & \frac{\partial p_3}{\partial y_k} & \frac{\partial p_3}{\partial \dot{x}_k} & \frac{\partial p_3}{\partial \dot{y}_k} \end{pmatrix} \tag{6}$$

In the application of Kalman filter, it is necessary to ensure that the driving noise and measurement noise of the system must be white noise. In the process, the driving noise and measurement noise of the system are colored, and the change of the actual environment leads to the change of noise. Finally, the deviation between the estimated value and the real value becomes larger and larger. At this time, the filter can not make the optimal estimation of the state. Therefore, this paper uses adaptive weighted Kalman filter to estimate the parameters of the model.

Adaptive Kalman filter mainly introduces fading factor to modify the error covariance matrix online and update the noise matrix and state noise matrix to prevent the divergence of the filter and improve the robustness of the algorithm. The adaptive Kalman filter algorithm introduces the weighting coefficient to adjust the measurement noise and state noise. The weighting coefficient is calculated as:

$$d_k = \frac{1 - b}{1 - b^{k+1}} \tag{7}$$

where k represents the current k time, b is forgetting factor, value takes from 0 to 1. It can be seen that when $k \to \infty$, d_k tends to 1, that is, the Kalman filter algorithm of the same standard.

The weighting coefficient is calculated and the noise is dynamically adjusted by the residual at the last time as follow:

$$\begin{cases} r_k = (1-d_k)r_{k-1} + d_k\left(Z_k - H_k\bar{X}_k^-\right) \\ R_k = (1-d_k)R_{k-1} + d_k\left(\varepsilon_k\varepsilon_k^T - H_kP_{k-1}H_k^T\right) \\ q_k = (1-d_k)Q_{k-1} + d_k\left(\bar{X}_k - \Phi_{k,k-1}\bar{X}_{k-1}\right) \\ Q_k = (1-d_k)Q_{k-1} + d_k\left(K_k\varepsilon_k\varepsilon_k^TK_k^T + P_k - \Phi_{k,k-1}P_k^-\Phi_{k,k-1}^T\right) \end{cases} \tag{8}$$

where $\varepsilon(t)$ represents the residual as follow:

$$\varepsilon(t) = Y_k - H_k\overline{X}_k^- \tag{9}$$

where, r_k and q_k respectively represent the mean value of measurement noise and state noise, R_k and Q_k represent the variance of measurement noise and state noise.

3.2 Combined Tracking with Camera Sensor After UWB Fusion

Vision-based target tracking algorithms cannot be completely reliable in real scenarios, and there is a possibility of misidentification in some extreme conditions. However, although UWB sensors have large errors in the measured data due to multipath effects, the fusion of IMU sensors can limit the errors to a certain range. Therefore, when the difference between the visually tracked target and the UWB fused target is small, the position of the visually tracked target can be considered more accurate. And when the difference between the position of the tracked target after UWB fusion and the position of the visually tracked target is large, it may be due to the misidentification of the visual tracking, so the position of the tracked target after UWB fusion is more reliable at this time, so the position of the UWB tracked target is mainly used at this time.

In this paper, three characteristics are used to determine which primary tracking method is used to drive the robot to achieve following. These include whether there is a tracked target in the camera field of view and whether the position of the visually tracked target point and the position of the UWB tracking point are on the same side of the x-axis of the camera coordinate system. The fusion status can be obtained as shown in Table 1.

Table 1. Fusion status.

Features	Target in camera field of view A_1	On the same side of x-axis A_2	UWB coordinates are within the camera range A_3	Primary tracking methods
1	Yes	Yes	Yes	Visual tracking position
2	Yes	Yes	No	Visual tracking position

(*continued*)

Table 1. (*continued*)

Features	Target in camera field of view A_1	On the same side of x-axis A_2	UWB coordinates are within the camera range A_3	Primary tracking methods
3	Yes	No	Yes	Visual tracking position
4	Yes	No	No	UWB tracking position
5	No	No	Yes	UWB tracking position
6	No	No	No	UWB tracking position
7	No	Yes	Yes	UWB tracking position
8	No	Yes	No	UWB tracking position

The decision tree ID3 algorithm uses the information gain criterion to select features for classification on each node of the decision tree, calculates the information gain for all possible features, and selects the feature with the greatest information gain as the classification criterion.

As shown in the above table, A_1 indicates that there is a target in the camera field of view, A_2 indicates that it is on the same side of the x-axis, and A_3 indicates that the coordinates after UWB fusion are on the same side of the camera. Suppose there are k categories in data set D that can be classified, and these categories can be represented by A_i(i = 1,…, k). The information gain of feature A to training data set D is expressed as $g(D \mid A)$ as shown below:

$$g(D \mid A) = H(D) - H(D \mid A) \tag{10}$$

It is defined as the difference between the empirical entropy $H(D)$ of set D and the empirical entropy of set D under the given characteristic A_k condition.

Empirical entropy $H(D)$ represents the uncertainty in set D, as shown below:

$$H(D) = -\sum_{k=1}^{K} \frac{|A_k|}{|D|} log\left(\frac{|A_k|}{|D|}\right) \tag{11}$$

The empirical entropy of set D under the given characteristic A_k condition as shown below:

$$H(D \mid A) = \sum_{k=1}^{n} \frac{|D_k|}{|D|} H(D_i) = -\sum_{i=1}^{n} \frac{|D_i|}{|D|} \sum_{k=1}^{K} \frac{|D_{ik}|}{|D_i|} log\left(\frac{|D_{ik}|}{|D_i|}\right) \tag{12}$$

The classification table of sensor fusion analyzed above in this paper is shown in Table 1. Then the fusion of UWB data and camera data is realized according to the ID3

algorithm of decision tree. According to Eq. (10), Eq. (11) and Eq. (12), the information gain based on different characteristics can be calculated, respectively, $g(D, A_1) = 0.548$, $g(D, A_2) = 0.0488$, $g(D, A_3) = 0.0488$, $g(D, A_1) > g(D, A_2) = g(D, A_3)$. Therefore, first selecting features for classification can reduce the uncertainty of class information more. According to $g(D, A_2) = g(D, A_3)$, it shows that selecting features A_2 and A_3 has the same effect in reducing the uncertainty of the set. Therefore, the decision tree model shown in Fig. 3 can be obtained. According to the decision tree model shown in Fig. 3, the fusion of UWB tracking data and camera tracking data after fusion with IMU can be realized.

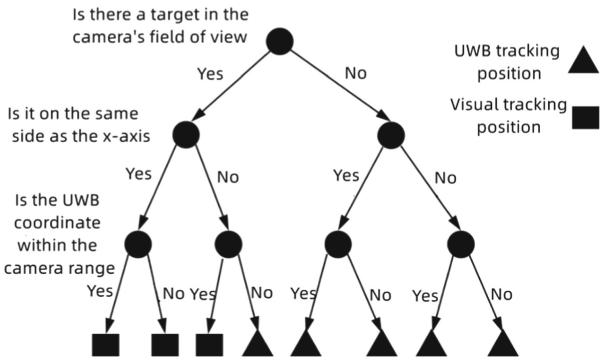

Fig. 3. Decision tree model of multi-sensor and camera fusion

4 Experiment Verification

This paper designs relevant verification experiments according to the design indexes of the following robot designed in this paper.

4.1 Following Distance Test

In order to verify the following distance, the experiment of following distance is carried out in this paper. The following scene is shown in Fig. 4, and the distance between the real-time following target and the following robot and the speed of the following car are recorded. As shown in Fig. 4 (a) and Fig. 4 (d), the robot can follow the target in a straight line or on a gentle slope; In Fig. 4 (b) and Fig. 4 (c), the robot follow the target to the right and the left respectively.

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

Fig. 4. Follow distance experiment

4.2 Following Occlusion Experiment

In order to verify the design index of the extraction distance of the occluded target, the tracking occlusion experiment is designed in this paper. The experiment verifies that the tracking vehicle can sense the tracking target within a certain range, and achieves ultra-broadband hyper-visual distance perception to extract the target position under visual target occlusion.

As shown in Fig. 5 (a), when moving to the position shown in Fig. 5 (b), the tracker can perceive the position of the target. When moving from the position shown in Fig. 5 (c) to the position shown in Fig. 5 (d), the visual tracker has failed and cannot perceive the position of the tracked target. As shown in the visual index diagram in Fig. 6, since the target is no longer within the camera field of view, there is no pixel error and overlap rate with the calibration frame, so they are all 0. As can be seen from Fig. 7, although the tracking accuracy of UWB is not high, it can also sense the position of the target. Therefore, it can be concluded from this experiment that when the tracking target is completely blocked, the following robot can effectively perceive the tracking target position.

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

Fig. 5. Human target tracking experiment

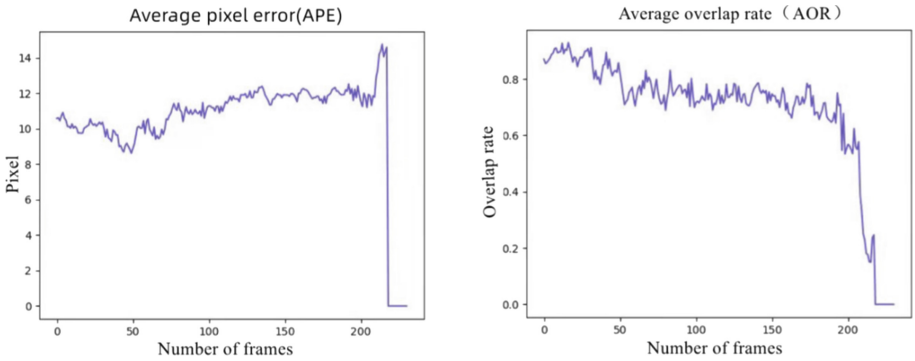

Fig. 6. Visual tracking index

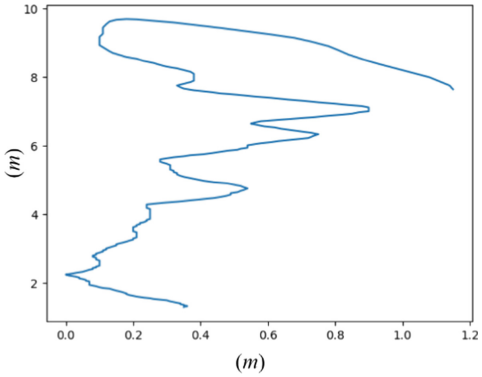

Fig. 7. Sage-huse adaptive Kalman filter fusion trajectory

4.3 PTZ Following Experiment

As shown in Fig. 8 (a) and (b), when the target leaves the camera field of view, the visual tracking algorithm fails and the target position cannot be perceived. As shown in Fig. 8 (c), when the target reappears within the camera field of view, the target tracking algorithm in this paper can continue tracking again. When moving from Fig. 8 (c) to Fig. 8 (d), the camera pan tilt can follow the target to ensure that the target is within the camera field of view. When moving the state of Fig. 8 (e), the target is partially obscured at this time, but the tracker can still track the target. After manually calibrating 1290 pictures, calculate the average pixel error (APE) and average overlap (AOR) with the tracking frame output by the tracking algorithm, as shown in Fig. 9.

APE is the error value based on the pixel distance between the predicted target center position and the real position, and the final result is averaged. AOR is the intersection ratio of the predicted area to the real area for each frame, and the final result is averaged.

As shown in Fig. 9 (a), there are four average errors of 0, which are in frames 85–101, 303–319, 837–900 and 1073–1092 respectively. This is because the target is completely obscured by obstacles in these frames. From the index of overlap rate, the overlap rate of

tracking target frame and calibration frame can be maintained above 0.6, which shows that when the target is within the field of view, the tracking algorithm can track the target and has a certain accuracy.

In conclusion, this experiment can verify that the tracking effect of the lightweight tracking network in this paper can meet the tracking requirements, and the two degree of freedom camera platform can realize the continuous tracking of the target.

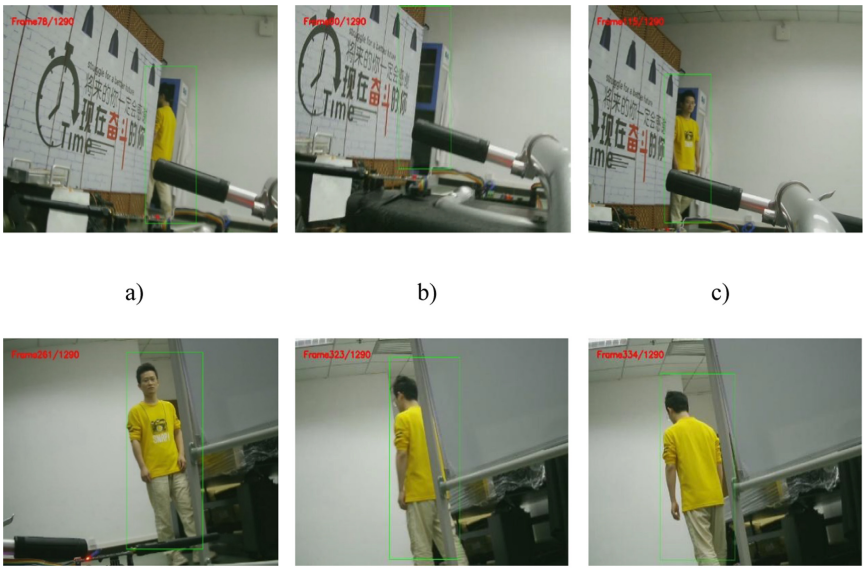

Fig. 8. Tracking of camera PTZ

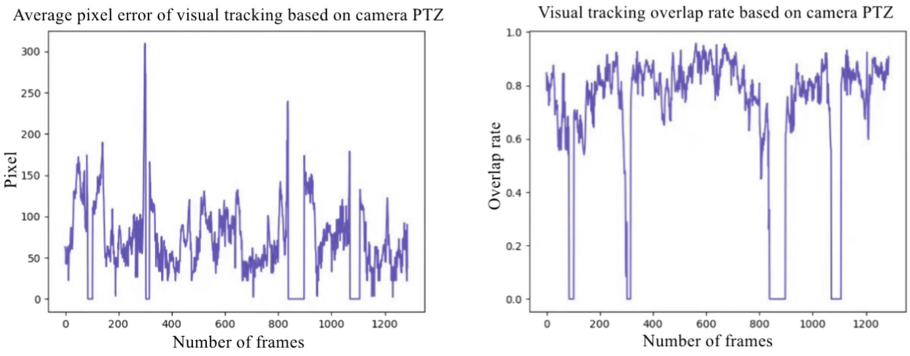

Fig. 9. APE and AOR indexes in camera PTZ following experiment

5 Conclusion

Taking the following robot as the application scenario, this paper designs the overall scheme of the following robot from the perspective of the reliability of the following robot, focuses on the research of human target tracking method based on multi-sensor, and the main results are summarized as follows:

(1) Aiming at the human target tracking method studied in this paper, a set of target following robot system is designed, including sensor type selection and so on. At the same time, the motion model of the robot is modeled to realize the motion control of the robot.
(2) Based on Kalman filter and decision tree, a tracking method combining UWB, IMU and monocular camera is proposed to realize that when the target is completely blocked by obstacles, the following robot can still perceive the problem of following the target so as to realize robust tracking.

References

1. Hirose, N., Tajima, R., Sukigara, K.: Personal robot assisting transportation to support active human life — human-following method based on model predictive control for adjacency without collision. In: 2015 IEEE International Conference on Mechatronics (ICM), pp. 76–81 (2015)
2. Kim, H., Lee, J., Lee, S., Cui, X., Kim, H.: Sensor fusion-based human tracking using particle filter and data mapping analysis in in/outdoor environment. In: 2013 10th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pp. 741–744 (2013)
3. Jung, E., Yi, B.: Study on intelligent human tracking algorithms with application to omnidirectional service robots. In: 2013 10th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pp. 80–81 (2013)
4. Wan, M., Zhang, H., Fu, M., Zhou, W.: Motion control strategy for redundant visual tracking mechanism of a humanoid robot. In: 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), pp. 156–159 (2016)
5. Jia, S., Zang, R., Li, X., Zhang, X., Li, M.: Monocular robot tracking scheme based on fully-convolutional siamese networks. In: 2018 Chinese Automation Congress (CAC), pp. 2616–2620 (2018)
6. Hare, S., Golodetz, S., Saffari, A., et al.: Struck: structured output tracking with kernels. IEEE Trans. Pattern Anal. Mach. Intell. **38** 2096–2109 (2015)
7. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. IEEE Trans. Pattern Anal. Mach. Intell. **6**(1), 1409–1422 (2010)
8. Zhang, K., Zhang, L., Yang, M.: Fast compressive tracking. IEEE Trans. Pattern Anal. Mach. Intell. **36**(10), 2002–2015 (2014)
9. Zhong, W., Lu, H., Yang, M.H.: Robust object tracking via sparse collaborative appearance model. IEEE Trans. Image Process. **23**(5), 2356–2368 (2014)
10. Danelljan, M., Bhat, G., Khan, F.S., et al.: ATOM: tracking by overlap maximization. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2020)

11. Ferrer, G., Zulueta, A.G., Cotarelo, F.H., et al.: Robot social-aware navigation framework to accompany people walking side-by-side. Auton. Robots **41**(4), 775–793 (2017)
12. Cifuentes, C.A., Frizera, A., Carelli, R., et al.: Human–robot interaction based on wearable IMU sensor and laser range finder. Robot. Auton. Syst. **62**(10), 1425–1439 (2014)
13. Jiang, B., Luo, R., Mao, J., et al.: Acquisition of localization confidence for accurate object detection (2018)