Active Tracking of Moving Objects with Speed Variations Using a Novel PTZ Camera-based Machine Vision Technique

Feng Lu[1], Youchun Xu[1], Yuan Zhu[1], Zhichao Zhang[1], Yulin Ma[2,*], Kang Mao[3]
[1]Institute of Military Transportation, Army Military Transportation University, Dongjuzi 1, Tianjin 300161, China
[2]School of Mechanical and Automotive Engineering, Anhui Polytechnic University, Beijing Mid Rd 1, Wuhu 241000, China
[3]Unit 32521, Zhenjiang, Jiangsu 212000, China
1849048346@qq.com; xu56419@126.com; mm1849048346@126.com; zzcjake6688@163.com; [*]mayulin@mail.ahpu.edu.cn; maokang94@163.com

***Abstract*—Most existing methods of active tracking focus mainly on slow-moving objects, resulting in limited adaptability to objects with variable speeds. To bridge this research gap, a novel spherical coordinate guided adaptive active tracking (SCAAT) approach based on the pan–tilt–zoom (PTZ) camera machine vision system is proposed in this study. For object detection and tracking, YOLOv5 and DeepSORT are employed in the PTZ vision system. The spherical coordinates and angular speeds of the moving object can be acquired under the spherical coordinate system. For practical application, the start-time and start-angle delay of the PTZ cameras are calibrated, and a speed control equation is conducted in the spherical coordinate system to reduce rapid location deviation between the PTZ and the moving object. To adapt different speeds of the object and avoid camera shaking under different zooms, an adaptive tracking window is designed to keep the object within the camera field of view. Experimental testing has been performed to evaluate the proposed SCAAT method. The results indicate that the SCATT can not only expand the effective following distance and zoom of the PTZ camera, but also effectively improve the accuracy and stability of active tracking for the moving object with large speed variations.**

***Index Terms*—Machine vision; Spherical coordinate; Active tracking; Moving objects; Speed variations.**

I. INTRODUCTION

Currently, the rapid advancement of sensor technologies, such as Lidar, cameras, and radar, has greatly enhanced the autonomous capabilities of unmanned ground vehicles (UGVs). Among various sensing technologies, object tracking is very important as it provides differential motion characteristics of the object, such as the speed and acceleration that are not observable. Moreover, object tracking can predict future motion states and estimate the trajectory [1]. Object tracking can be classified into passive and active modes on the basis of the relative motion state between the sensor and the object. Passive tracking extracts data sequences from environmental information collected with stationary/fixed sensors. Active tracking uses active sensors, such as pan-tilt-zoom (PTZ) cameras, to dynamically follow and record the object.

Among the various passive tracking methods, simple online and real-time tracking (SORT) [2] has been widely used due to its excellent performance at high frame rates. The SORT utilises a simple Kalman filter to address interframe data correlation and employs the Hungarian algorithm for correlation measurement. Furthermore, SORT with a deep association metric (DeepSORT) uses the convolutional neural networks (CNN) to extract features of the tracking object, thereby improving tracking accuracy [3]. Since passive tracking can only extract object information within a limited field of view (FOV) of a camera, multiple cameras are often used to detect and track the objects. However, passive tracking may lead to unsatisfied tracking, as the object may often disappear in one camera and appear in the other. To address this issue, scholars have delved into object reidentification technology. The fast-reid module has been proposed to enhance the robustness of reidentification through the modular encapsulation and integration of various techniques in multiple baselines [4]. Liu *et al.* [5] proposed the tracklet filter strategy (TFS) in adjacent cameras to effectively cluster object trajectories. Wu, Qian, Wang, and Yang [6] employed a template matching method using EfficientDet as a detector and DeepSORT as a tracker to reconnect discontinuous trajectories. However, these approaches increase the time consumption and complexity of the solution.

In contrast, the active sensor can adjust its view to follow the moving object, which avoids the aforementioned problem of multiple sensors in passive tracking. As one of the most popular active sensors, the PTZ camera can keep the moving object within its FOV, making it possible to use all computational resources on the relevant element of the scene [7]. As a result, the PTZ camera is the popular choice for

Manuscript received 3 January, 2025; accepted 22 February, 2025.

active tracking [8]. Various handheld pan-tilt tracking devices are provided for mobile phones with rotation and zoom functions [9] to enhance the capacity and flexibility of PTZ camera-based tracking. Muñoz, Ruiz-Santaquiteria, Deniz, and Bueno [10] introduced a rule-based approach to enhance the tracking and detection of potential firearms.

Additionally, reinforcement learning is utilised to train the active trackers to achieve optimal alignment between the perceived object and the ideal camera position. For example, Zhong, Sun, Luo, Yan, and Wang [11] adopted an asymmetric dueling mechanism to train the tracker based on adversarial reinforcement learning. The tracker and object were trained in a duel/competition manner, where the tracker aimed to lock on the object while the object attempted to evade it. Luo *et al.* [12] integrated convolutional neural network (CNN), long short-term memory, and actor-critic network to guide the behaviour of the tracker and enable an accurate estimation of its state value without rotational interference. Fang, Liu, Wen, Yang, Li, and Han [13] used integrated both discrete and continuous action spaces for deep reinforcement learning to enhance the generalisation ability of agent decisions across diverse scenarios. However, these active tracking methods based on reinforcement learning are primarily evaluated in virtual environments. There exists a substantial gap between simulated active trackers and practical requirements. To bridge this research gap, this study aims to utilise the directionality and zoom capabilities of the PTZ cameras for the moving object tracking with obvious speed variations. Despite the significant success achieved by current active tracking methods, they are limited in their ability to track objects with speed variations because of reliance on image coordinates control. In addition, their performance may not be guaranteed in real-world application scenarios, such as time and angle delays of the camera motion and various object speed with focal length changes. Moreover, most existing algorithms focus on enhancing the tracking stability by fixed tracking window settings, but ignore the disadvantages of the fixed tracking window.

In this paper, a novel SCAAT method is proposed to solve the aforementioned problems in active tracking. The primary contributions include the following:

1. The PTZ camera is guided to actively track the object under the spherical coordinate system with the feature of adaptive adjustment of the camera's FOV;
2. The start-time and start-angle delay of the PTZ camera are calibrated. The position and speed of the moving object are integrated to develop a speed control equation to enhance the speed control accuracy of the PTZ cameras;
3. An adaptive tracking window is designed to improve the stability of the active tracking, which allows one to automatically adjust the tracking window size according to the actual state of the PTZ camera and the moving object.

The structure of this paper is organised as follows. In Section II, an overview of current technology and challenges of active object tracking using the PTZ camera are discussed. In Section III, YOLOv5 and DeepSORT are introduced to realise object detection and tracking. Section IV presents a performance evaluation of object tracking using the PTZ camera. Section V concludes this study.

II. RELATED WORK

The PTZ camera-based active tracking usually includes object tracking and camera control [14]. In addition, some scholars have conducted research on mirror-driven pan-tilt cameras. Subsequently, we state the main difference between our method and the others.

A. Object Detection and Tracking

Object detection and tracking using PTZ cameras have been extensively studied in recent years. Yun, Kim, Bae, and Park [15] proposed a method for detecting moving obstacles using two background models, where a wide background model was used to detect significant changes in the environment, while the other model was used to detect minor changes. However, deficient feature information between the two models would lead to a failed detection for multiple environments. Zhu, Yang, and Yu [16] proposed a tracking method based on YOLOv4 and Hu, Zhang, Xu, and Chen [17] used YOLOv5 using multiple PTZ cameras. The position of the object was determined by rotating the PTZ cameras. Furthermore, different coordinate systems have been explored to locate objects [18], [19]. The image coordinate system is a popular selection for location information extraction because of the easy and fast control of the camera movement. However, it still remains a challenge that the history information is difficult to fuse to control the PTZ cameras because of multiple zooms.

B. Perception-Guided Control

In active tracking, the camera controller adjusts the yaw and pitch angles to keep the object in camera view and optimises the zoom value to maximise the object visibility [20]. The proportional-integral-derivative (PID) control is often adopted with a target tracking window as the camera controller [21]–[23]. Li, Pan, and Cheng [23] and Ji [24] used a fixed size of the target tracking window to control the PTZ camera. Li, Li, Chen, and Zhao [25] established a linear relationship between the PTZ camera's control angle and the centre offset to maintain active tracking. However, it still remains an unsolved problem that the fixed tracking window cannot process speed variations of moving objects at different zooms.

C. Mirror-Driven Pan-Tilt Cameras

In addition to mechanical PTZ cameras, mirror-driven pan-tilt cameras are also extensively investigated. These types of cameras can capture moving objects through ultra-fast gaze control of the camera rotation and zooming. Hu, Shimasaki, Jiang, Senoo, and Ishii [26] controlled the rotation angles of multiple pan-tilt cameras. Shen, Hu, Shimasaki, Senoo, and Ishii [27] developed a multiface zoom tracking system to detect multiple pedestrians. However, the mirror-driven camera technique is limited to capturing multiposition and multifocal length within a fixed FOV. In contrast, the mechanical PTZ cameras can provide 360-degree coverage possess irreplaceable advantages.

To address the unresolved problems mentioned in the related work, a new SCAAT method is proposed to track moving objects with various speeds, where the relationship between the tracking window, the camera motion, and the

object speed is considered to enhance the stability of the active tracking.

III. PROPOSED ACTIVE TRACKING METHOD

A. Spherical Coordinate System

The PTZ cameras mainly capture scene information around the rotation centre. Therefore, a spherical coordinate system (SCS) is established based on the rotation centre, which assumes that the optical centre of the camera coincides with the rotation centre. Figure 1 illustrates the established SCS.

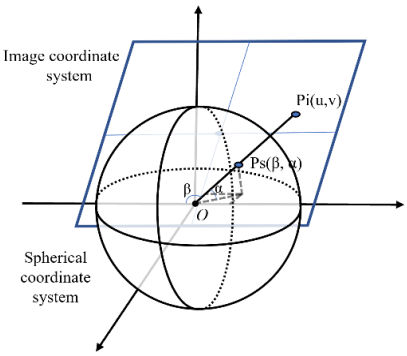


Fig. 1. Spherical coordinate system (SCS).

The conversion of the image coordinate system (ICS) to the camera coordinate system can be well described as follows:

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}, \quad (1)$$

where (f_x, f_y) is the focal length, (c_x, c_y) is the position of the principal point of the camera, $(u, v)^T$ is the pixel coordinate, and $(x_1 \;\; y_1 \;\; z_1)^T$ is the homogeneous coordinate of the object point in the camera coordinate system $(z_1 = 1)$. The coordinates under the camera coordinate system are normalized like $(x_2, y_2, y_2)^T$ is given by:

$$\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \frac{1}{\sqrt{\left(x_1^2 + y_1^2 + z_1^2\right)}} \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}. \quad (2)$$

The spherical coordinate can be transformed from the normal camera coordinate:

$$\begin{pmatrix} \beta \\ \alpha \end{pmatrix} = \begin{pmatrix} \pi - arccosx_2 \\ arctan\dfrac{y_2}{z_2} \end{pmatrix}. \quad (3)$$

The point in the SCS at the time of *t* can be formulated as:

$$\begin{pmatrix} \beta_w(t) \\ \alpha_w(t) \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & p(t)-p(0) \\ 0 & 1 & t(t)-t(0) \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_c(t) \\ \alpha_c(t) \\ 1 \end{pmatrix}, \quad (4)$$

where $(\beta_w(t),\ \alpha_w(t))^T$ is the world spherical coordinate of the point at the time of *t*, $(p(0), t(0))$ and $(p(t), t(t))$ are the pose of the PTZ camera at $t = 0$, and $t = t$,$(\beta_c(t), \alpha_c(t))^T$ is the camera spherical coordinate of the point.

B. Object Detection and Tracking

Object detection. For object detection, the two-stage R-CNN series [28], [29] and single-stage detectors like RetinaNet [30] and YOLO [31] are among the most popular frameworks. YOLOv5 [32] is one of the most advanced object detection methods in YOLO series. The feature extraction network of YOLOv5 primarily employs multiple residual components to extract image features and generates a multilayer subsampling feature. The predictive network utilises the convolution module to achieve confidence scores and regression coefficients at different sampling multiples by using results from various layers.

Object tracking. Most existing multiple object tracking (MOT) methods can be grouped into [32] detection-based tracking (DBT) and detection-free tracking (DFT) [33]. DBT is more suitable for the PTZ camera [34] and DeepSORT is able to track moving objects through longer periods of occlusions and effectively reduce the number of identity switches.

As a result, this study adopts YOLOv5 and DeepSORT for object detection and tracking.

C. Camera Controller

To deal with various speeds of the object, the PTZ camera is controlled using the PID by comprehensively considering both the angular speed of the object and the deviation from the image centre. The delay calibrating and compensating system is also considered in the PID controller. The delay includes the start-angle delay and the start-time delay. The start-angle delay is described as the delay from stop status to movement; while the start-time delay is described as the angle generated by the PTZ inertia during the switch between start and stop control.

When the deviation between the object and the image centre becomes larger, the camera movement should speed up to keep up with the object. At the same time, due to the system delay, a larger speed control value and short control time are required to reduce the gap between the image centre and the object centre. Initially, the angular speed of the object can be calculated by:

$$\begin{pmatrix} \omega_{\beta w}(t) \\ \omega_{\alpha w}(t) \end{pmatrix} = \frac{1}{T(t)-T(t-1)} \begin{pmatrix} \beta_w(t)-\beta_w(t-1) \\ \alpha_w(t)-\alpha_w(t-1) \end{pmatrix}, \quad (5)$$

where $T(t), T(t-1)$ are the timestamp between *t* and $t-1$, $\beta_w(t), \beta_w(t-1)$ are the latitude coordinates, and $\alpha_w(t), \alpha_w(t-1)$ are the longitude coordinates. The motion diagram is shown in Fig. 2.

Considering the angular speed of the object and the distance deviation from the centre of the object to the centre of the image, the PTZ camera will adjust itself to ensure that the object stays within its FOV. To achieve this goal, the error equation is established as:

$$\begin{pmatrix} er_\beta(t) \\ er_\alpha(t) \end{pmatrix} = \begin{pmatrix} \left|\beta_c(t)-\beta_{cc}(t)\right| \\ \left|\alpha_c(t)-\alpha_{cc}(t)\right| \end{pmatrix}, \quad (6)$$

where $(\beta(t), \alpha(t))$is the longitude and latitude of the object, while the image centre under SCS is $(\beta_{cc}, \alpha_{cc})$. Assuming that the direction of the image centre to the object centre in the image is similar to the direction of the object's motion, the rotation and pitch speed of the PTZ camera are required to satisfy the equation as follows:

$$\begin{pmatrix} \omega_{\beta c}(t)\times\Delta t \\ \omega_{\alpha c}(t)\times\Delta t \end{pmatrix} = \begin{pmatrix} er_{\beta}(t)+\omega_{\beta w}(t)\times(\Delta t+\tau_{\beta i})-\theta_{\beta i} \\ er_{\alpha}(t)+\omega_{\alpha w}(t)\times(\Delta t+\tau_{\alpha i})-\theta_{\alpha i} \end{pmatrix}, \quad (7)$$

where, $\triangle t$ is the actual rotation time of the PTZ camera. $\tau_{\beta i}$,$\tau_{\alpha i}$ is the start time delay during rotation and pitch, at the gear of i $(1 \le i \le N)$, can be described as

$$\tau_{\delta} = \begin{cases} \tau_{\delta i}, \text{ start - time delay at gear } i, \delta \text{ is } \alpha \text{ or } \beta, \\ 0, \qquad\qquad \text{gear switch,} \end{cases} \quad (8)$$

where $\theta_{\beta i}$,$\theta_{\alpha i}$ is the start-angle delay during rotation and pitch, at the gear of i $(1 \le i \le N)$, can be described as

$$\theta_{\delta} = \begin{cases} \theta_{\delta i}, \text{ start - angle delay at gear } i, \delta \text{ is } \alpha \text{ or } \beta, \\ 0, \qquad\qquad \text{gear switch.} \end{cases} \quad (9)$$

The start-time delay τ_{δ} and start-angle delay θ_{δ} can be estimated in Section IV-C.

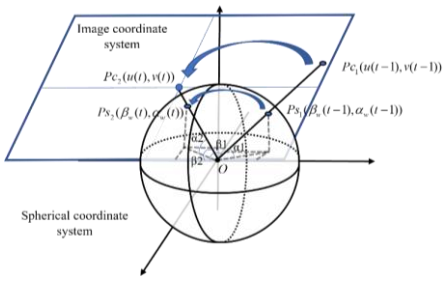


Fig. 2. Point motion and transformation in SCS.

To track the object, (7) must be met where $\triangle t \ge 0$, that is:

$$\begin{cases} \omega_{\beta c}(t) \ge \omega_{\beta}, \\ \omega_{\alpha c}(t) \ge \omega_{\alpha}. \end{cases} \quad (10)$$

The time delay τ_{δ} from starting to movement can be estimated at different speed gears (Section IV-C). Incremental PID is established as follows

$$\begin{aligned} \Delta\omega(t) = k_p\left(er(t)-er(t-1)\right)+k_i er(t)+ \\ +k_d\left(er(t)-2er(t-1)+er(t-2)\right), \end{aligned} \quad (11)$$

where $\omega_{\beta c}(t)$, $\omega_{\alpha c}(t)$ can be represented as:

$$\begin{cases} \omega_{\beta c}(t) = \Delta\omega_{\beta}(t)+\omega_{\beta w}(t), \\ \omega_{\alpha c}(t) = \Delta\omega_{\alpha}(t)+\omega_{\alpha w}(t). \end{cases} \quad (12)$$

Since the PTZ camera has N gears speed, we can select the gear i, j $(1 \le i, j \le N)$ for β, α, respectively,

$$\begin{cases} \omega_{\beta(i-1)} < \omega_{\beta c}(t) \le \omega_{\beta i}, \\ \omega_{\alpha(j-1)} < \omega_{\alpha c}(t) \le \omega_{\alpha j}, \end{cases} \quad (13)$$

where $\omega_{\beta 0}, \omega_{\alpha 0}$ are set to 0.

D. Adaptive Tracking Window

The tracking window is usually set at fixed size [23]–[25], in which it is considered a static area. The outer area of the tracking window is considered as the rotation and pitch area. However, its limitations in variable zooms are noteworthy. Assume in a scenario that the object moves in the same direction with the PTZ camera. When the tracking window size is large, the camera can achieve relatively stable movement, yet minimal zoom adjustment. The active following distance between the camera and the object is limited. Furthermore, the object is prone to move out of the camera's FOV once it accelerates. When the tracking window is small, the zoom can increase more, so that the object far from the camera can be kept within the camera's FOV. This leads to a higher tolerance for the object speed, but can also cause the object to sway around the tracking window easily, resulting in PTZ camera shaking.

The PTZ camera used in this paper has only seven gears speed available, each corresponding to a rough angular speed. The camera speed controller has also seven gears, which can be defined as vp_i, vt_i, t = 1, 2, …, 7. According to Section III-A, homogeneous coordinates are$(x(t), y(t))$. The local spherical coordinate is $(\beta_c(t), \alpha_c(t))$. After the camera rotates the angle $\triangle\beta(t)$, $\triangle\alpha(t)$, the local spherical coordinate of point A is $(\beta_c(t+1), \alpha_c(t+1))$. Since point A is homogeneous, $x_1, y_1, x_2, y_2 \ll 1$, the image coordinate of point A can be calculated as:

$$\begin{cases} u(t) \approx -f_x \times cos\beta_c(t)+c_x, \\ v(t) \approx f_y \times tan\alpha_c(t)+c_y. \end{cases} \quad (14)$$

The change of the image coordinate in u and v is formulated as:

$$\begin{cases} \Delta u(t) = f_x\left(cos\left(\beta_c(t+1)\right)-cos\beta_c(t)\right), \\ \Delta v(t) = f_y\left(tan\left(\alpha_c(t+1)\right)-tan\alpha_c(t)\right). \end{cases} \quad (15)$$

According to (4), the maximal distance between adjacent frames of the target movement in a spherical coordinate system is estimated as:

$$\begin{cases} \beta_c(t+1)-\beta_c(t) = \omega_{\beta w}(t)\tau_{\beta}-\theta_{\beta}, \\ \alpha_c(t+1)-\alpha_c(t) = \omega_{\alpha w}(t)\tau_{\alpha}-\theta_{\alpha}. \end{cases} \quad (16)$$

When the object departs from the tracking window after adjusting of yaw and pitch, the system generates a motion command for the PTZ camera to pull the object back within the tracking window. If the size of the tracking window is below the maximum distance between adjacent frames of the target's movement, the active tracking system may easily lead to oscillation. To deal with it, it is necessary to ensure that the object motion in ICS meets the following requirements:

$$\begin{cases} w \ge |\Delta u_{max}|, \\ h \ge |\Delta v_{max}|. \end{cases} \quad (17)$$

To maintain system stability, we limit the threshold sizes of the tracking window between (w_{min}, h_{min}) and (w_{max}, h_{max}), respectively, within which the size is dynamically adjusted based on the movement of the target. The adaptive adjustment formula of the tracking window is as follows:

$$w_{adaptive} = \begin{cases} w_{min}, |\Delta u_{max}| \le w_{min}, \\ |\Delta u_{max}|, w_{min} < |\Delta u_{max}| < w_{max}, \\ w_{max}, |\Delta u_{max}| \ge w_{max}, \end{cases} \quad (18)$$

$$h_{adaptive} = \begin{cases} h_{min}, |\Delta v_{max}| \le h_{min}, \\ |\Delta v_{max}|, h_{min} < |\Delta v_{max}| < h_{max}, \\ h_{max}, |\Delta v_{max}| \ge h_{max}. \end{cases} \quad (19)$$

Consequently, $(w_{adaptive}, h_{adaptive})$ is mainly determined by the object movement speed and current camera zoom.

It is worth noting that the effective value f_x, f_y can be estimated when $w = w_{max}$ or $h = h_{max}$. Subsequently, the maximum effective value $zoom_x, zoom_y$ can be obtained. We can achieve the maximum zoom of the system control for effective active tracking, as follows

$$zoom = \left(zoom_x, zoom_y \right). \quad (20)$$

If zoom exceeds the threshold, active tracking will be unstable.

The overview of the proposed SCAAT method for active tracking is illustrated in Fig. 3, where the whole model is mainly composed of the spherical coordinate system, object detection, and tracking and perception-guided control modulus.

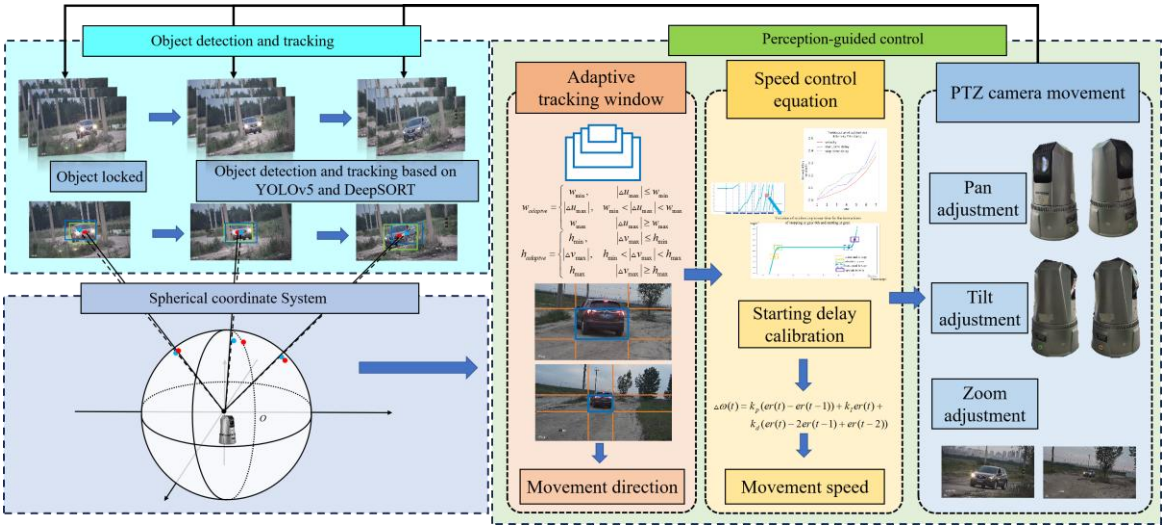


Fig. 3. Overview of the proposed SCAAT method for active tracking.

IV. EXPERIMENTAL VALIDATION AND RESULTS

A. Experimental Setup

The proposed method is implemented in C++ on a PC with an Intel Core i5-7300HQ 2.5 GHz CPU and a NVIDIA GeForce GTX 1060 GPU. Software design consists primarily of four control threads to facilitate active tracking, consisting of the data acquisition module, the data reception and perception module, the perception-guided control module, and control command execution module.

The data acquisition module is responsible for obtaining and transmitting images, as well as camera parameters (i.e., the pan, tilt, and zoom). The data reception and perception module involves receiving the data set and applying YOLOv5 and DeepSORT to achieve object detection and tracking. The perception-guided control module includes the motion analysis based on speed control equation and tracking window, which generate control values for camera pan, tilt, and zoom. The control command execution module is responsible for delivering the control commands to the PTZ camera to maintain active tracking of the object. The system configuration parameters for the experiment verification are shown in Table I. Real-world data are collected using a PTZ camera to track with the movement objects of vehicles and persons.

TABLE I. PARAMETERS OF ACTIVE TRACKING SYSTEM.

	Parameter	Value
PTZ camera	Resolution	[1280, 720]
	Frame rate/FPS	25
	Zoom range	1–30
	Pan range/°	0-360
	Tilt range/°	-40–90
YOLOv5	Baseline	YOLOv5s
	Confidence	0.25
	IOU threshold	0.45
DeepSORT	Max distance	0.2
	Min confidence	0.3
	NMS max overlap	0.5
Tracking window size	Rect_min	[320, 180]
	Rect_max	[960, 540]

B. Evaluation Metrics

In real-world scenarios, it is necessary to perform active tracking to ensure the object is in the camera FOV. To properly assess the performance of the proposed active object tracking, the deviation between the object centre and the image centre, the following evaluation metrics are adopted:

$$TPO_w = \frac{1}{N} \sum_{t=1}^{N} TPO_{t,w}, \quad (21)$$

$$TPO_h = \frac{1}{N} \sum_{t=1}^{N} TPO_{t,h}, \quad (22)$$

where the TPO_w (Object Point Offset) and TPO_h (Image Point Offset) evaluate the quality of the camera control, and:

$$TPO_{t,w} = \frac{C_{FOV,w} - C_{PT,w}}{size_w}, \quad (23)$$

$$TPO_{t,h} = \frac{C_{FOV,h} - C_{PT,h}}{size_h}, \quad (24)$$

where the ($C_{PT,w}$, $C_{PT,h}$) is the object centre and the ($C_{FOV,w}$, $C_{FOV,h}$) is the image centre. ($Size_w, Size_h$) is the width and height of the image. One can note that the active tracking performance increases with decreasing TPO_w and TPO_h.

The effective centre proportion (ECP), which is the distribution probability within the tracking window for the coordinates of the centre of the object, is used to estimate the deviation between the centre of the object and the centre of the image. In general, a high ECP value indicates a low TPO condition.

Precision plots are visualisation chart that evaluates the spatial accuracy of a tracker by measuring the proximity of the predicted centre of bounding boxes to the image centre. Specifically, it plots the precision score against different distance thresholds across a sequence of frames.

The duration of active tracking is used to demonstrate the effectiveness of active tracking in real-world environments. A longer tracking duration indicates better active tracking performance.

The effective distance between the object and the PTZ camera is used to evaluate the active tracking performance, and the effective zoom is used to examine the max effective zoom for full use of the PTZ cameras. In general, a higher value of the effective zoom indicates a longer effective distance.

Stable movement of the PTZ cameras is essential for active tracking. The stability ratio is used to evaluate the stable movement of the PTZ cameras in this study. The stability ratio is defined as the proportion of the object moving towards the controlled direction. The higher the stability ratio, the more stable the active tracking.

C. Experimental Results

Delay calibration. By means of start-stop control, the camera gear is shifted every 20 seconds, followed by a 5-second pause to measure the mechanism delay of the start-time and start-angle. Meanwhile, the PTZ parameters are recorded at different moments. The test is carried out across seven gears to calculate the angular speed, start-time, and start-angle delay at different gears. For example, the change of the rotation angle (pan) is shown in Fig. 4, and Fig. 5 zooms the 6th gear curve in Fig. 4. One can estimate the start-time delay τ_β and start-angle delay $\triangle\beta$ between the initiate PTZ motion and the actual motion for different speed gears.

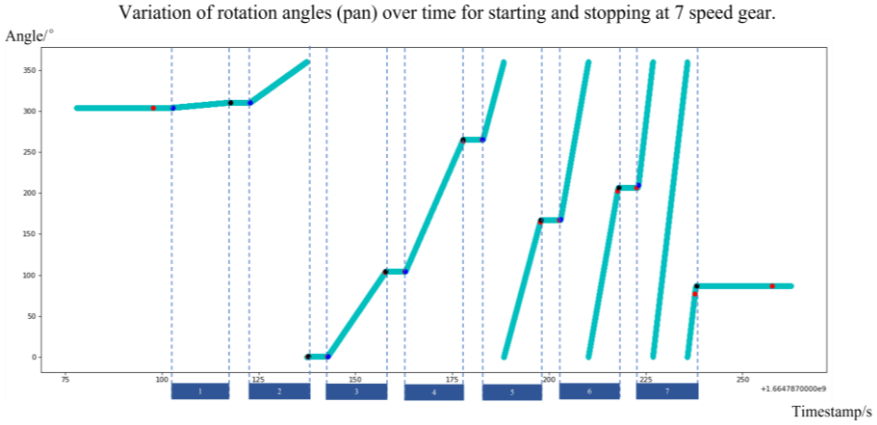


Fig. 4. Rotation changes under start-stop control commands.

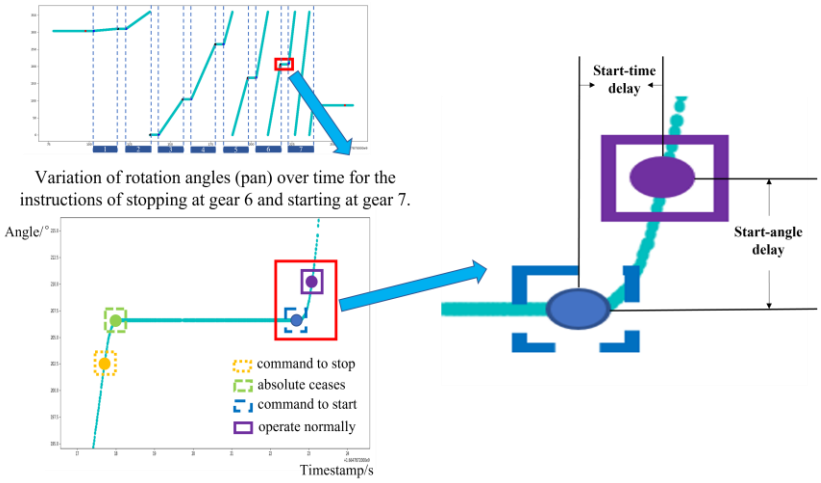


Fig. 5. Rotation changes under start-stop control commands in the 6th gear.

The relationship between the yaw angular speed, start-time, and start-angle delay with seven gears in the rotation

direction is shown in Figs. 6 and 7. As can be seen in the figures, the start-time delay and start-angle delay can be estimated from the rotation change curves and the delay values increase with the gear number.

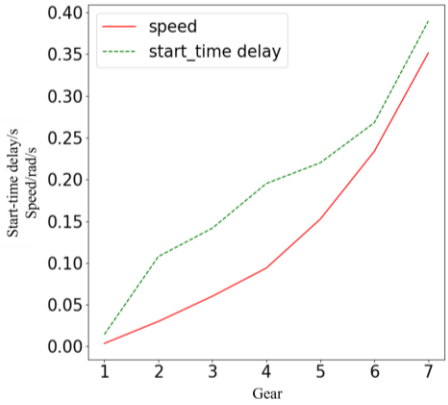


Fig. 6. Variational speed and start-time delay with seven-level gears in rotation direction.

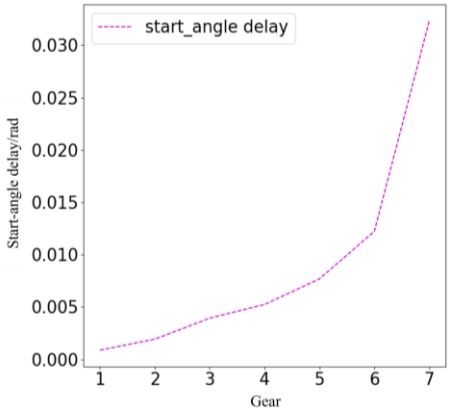


Fig. 7. Variational start-angle delay with seven-level gear.

Spherical coordinate system (SCS). The proposed active tracking method is implemented in the SCS. To highlight the superiority of the SCS, the performance of the proposed object tracker is compared between the SCS and the popular ICS [25]. To ensure the same experimental setting, the fixed tracking window is used for both the SCS and ICS in the comparison. The coordinates of the fixed tracking window are set from point [480, 270] to [800, 450]. To make the object centre remain within the tracking window, the x-coordinate of the object centre should be in the range of [480, 800], while the y-coordinate should be in the range of [270, 450]. The comparison results are depicted in Fig. 8, where the effective centre proportion in the x-coordinate using the SCS is 79.29 %, increased by 6.72 % compared with that using the ICS; while the effective centre proportion in the y-coordinate is using the SCS 94.66 %, increased by 3.76 % compared with the ICS.

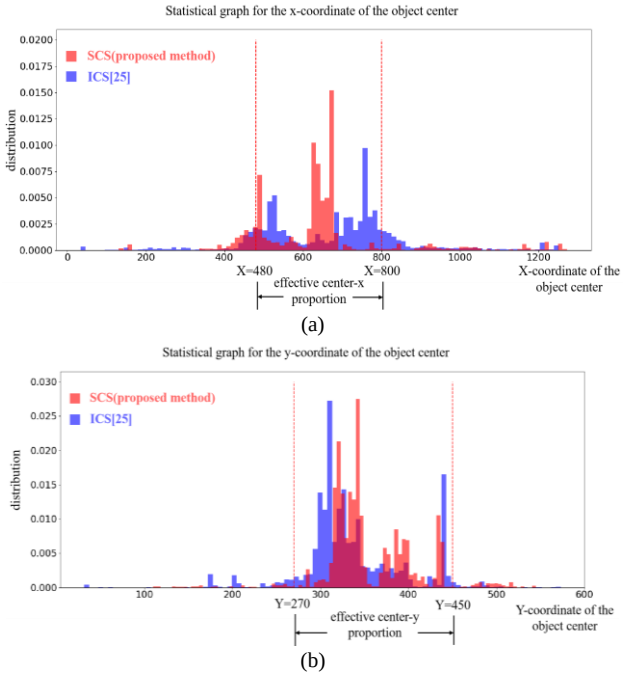


Fig. 8. Comparison results of the effective centre proportion: (a) X-coordinate and (b) Y-coordinate.

Camera control delay. To effectively control the camera movement, start-time and start-angle delay of mechanical motion are taken into account to improve the active tracking performance in real-world experiment tests. By considering the start-time and start-angle delay, the controller can select a suitable gear to make the PTZ camera quickly approach the object and shorter the control time to meet the hardware requirements. Table II lists the effective distance and zoom, TPO_w and TPO_h of the active tracking results. As can be seen in the table, if considering both the start-time and start-angle delay, the tracking duration can be prolonged by 7.6 s and the effective distance can be expanded by 8.53 m when compared with that if not considering these two factors, while the TPO_w and TPO_h can be improved by 11.3 % and 5.72 %, respectively. One can also note that the proposed method produces better results than these if one only considers one factor.

TABLE II. CAMERA CONTROL DELAY AND ACTIVE TRACKING RESULTS.

Start-angle Delay	Start-time Delay	Tracking duration	Effective distance	TPO_w/%	TPO_h/%
Not consider	Not consider	30.8 s	35.32 m	25.67	16.47
Consider	Not consider	33.1 s	39.54 m	18.59	12.83
Not consider	Consider	35.5 s	38.68 m	17.55	11.42
Consider	Consider	38.4 s	43.85 m	14.37	10.75

Adaptive tracking window. To assess the efficiency of the proposed adaptive tracking window, an experimental evaluation is conducted using different fixed tracking windows and the adaptive tracking window. The comparison results are illustrated in Figs. 9 and 10. The evaluation results indicate that the proposed adaptive tracking window is more effective in adapting to object movements. This is because the tracking window size can be adjusted on the basis of the current focal length, and the focal length and spherical coordinate can be further fine-tuned on the basis of the tracking window size.

Quantitative analysis. To comprehensively evaluate the performance of the proposed PTZ-based SCAAT, multiple experiments are carried out to track vehicles and persons with various speed variations.

We compared the proposed SCAAT method with recent

related approaches, including WD [10] and AVE [13], for active tracking of persons and vehicles under various conditions.

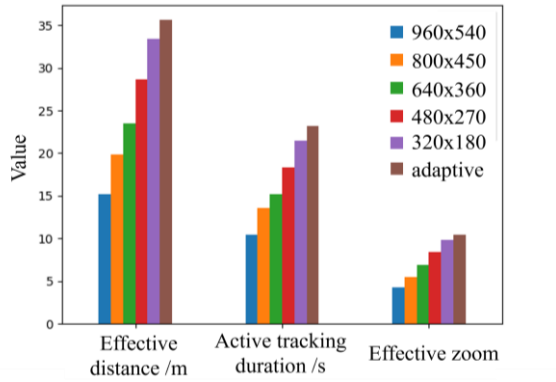


Fig. 9. Effective distance, active tracking duration, effective zoom using different tracking window settings.

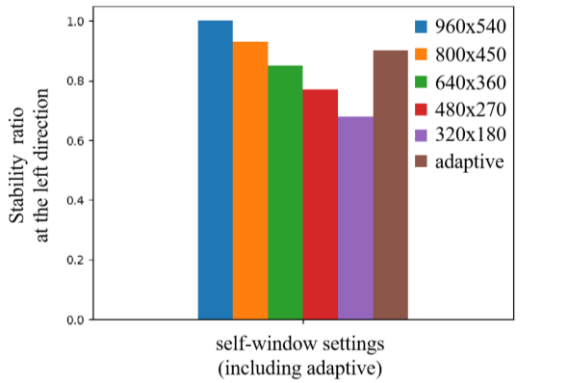


Fig. 10. Stability ratio using different tracking window settings.

The results, presented in Table III, indicate that although each method is capable of maintaining active object tracking, SCAAT, which incorporates spherical coordinates and control delays into the tracking process, can achieve an 8.98 % average increase in the effective distance, an 8.37 % improvement in the effective zoom, and a 1.62 % rise in the stability ratio. Furthermore, as illustrated in Fig. 11, the use of spherical coordinates alongside the inherent control delay of the PTZ camera enables a more accurate location of the tracked object within the field of view of the camera, resulting in a 2.4 % improvement in the tracking precision. Consequently, SCAAT exhibits robust performance under adversarial environmental conditions, outperforming existing methods in effective distance, focal length accuracy, stability ratio, and overall tracking precision.

In addition, we conduct experiments on the person and vehicle tracking under adversarial conditions of partial occlusions, rain, and smog. The temporal visual results of active tracking are shown in Fig. 12. The pan, tilt, and zoom values of the PTZ camera are constantly adjusted throughout the tracking process to ensure that the centre of the object remains within the tracking window. “Pan”, “Tilt”, and “Zoom” below the figure are described as the current camera parameters. Adjustment in the camera parameters of Fig. 12(a) can be checked in Fig. 13. Note that the SCAAT can maintain stable active tracking by adjusting the PTZ parameters over sequential frames across various scenarios.

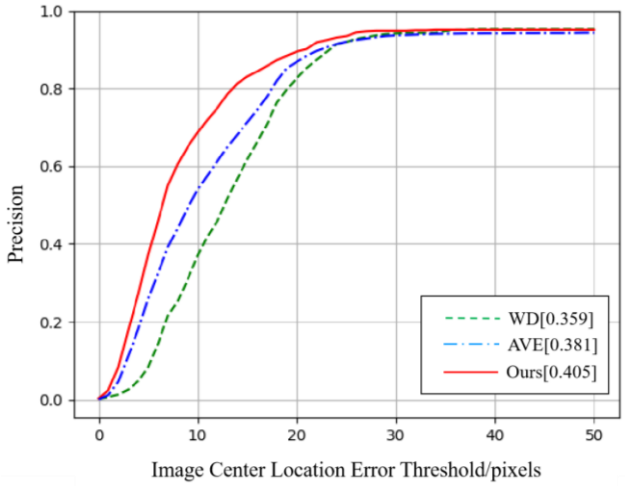


Fig. 11. Precision plots.

TABLE III. STATISTICAL TABLE OF EXPERIMENTAL RESULTS.

No.	Environment	Category	Method	Frame	Distance	Zoom	Stability ratio↑
1	sunny day	Person	WD [10]	2875	40.2 m	7.9	83.64 %
2		Person	AVE [13]	2764	45.8 m	8.3	85.26 %
3		Person	SCAAT	2949	**52.3 m**	**8.7**	**86.16 %**
4	rain	Person	WD [10]	2968	38.6 m	7.3	81.35 %
5		Person	AVE [13]	3021	42.5 m	7.6	83.59 %
6		Person	SCAAT	3156	**50.7 m**	**8.6**	**85.67 %**
7	smog	Person	WD [10]	3057	32.5 m	6.6	77.49 %
8		Person	AVE [13]	3091	36.7 m	7.1	79.56%
9		Person	SCAAT	3117	**40.2 m**	**7.9**	**82.18 %**
10	partially occlusion	Person	WD [10]	2986	33.7 m	6.8	76.52 %
11		Person	AVE [13]	3025	36.2 m	7.1	78.16 %
12		Person	SCAAT	3184	**39.4 m**	**7.7**	**80.07 %**
13	sunny day	Vehicle	WD [10]	5648	73.5 m	9.2	76.59 %
14		Vehicle	AVE [13]	5692	76.1 m	9.8	78.47 %
15		Vehicle	SCAAT	5766	**78.7 m**	**10.5**	**80.50 %**
16	rain	Vehicle	WD [10]	5269	67.5 m	7.8	83.48 %
17		Vehicle	AVE [13]	5316	72.3 m	8.7	84.69 %
18		Vehicle	SCAAT	5394	**76.9 m**	**9.6**	**85.24 %**
19	smog	Vehicle	WD [10]	5167	67.3 m	7.8	80.64 %
20		Vehicle	AVE [13]	5232	70.5 m	8.9	82.45 %
21		Vehicle	SCAAT	5304	**73.8 m**	**9.2**	**83.56 %**
22	partially occlusion	Vehicle	WD [10]	5258	71.6 m	8.9	76.34 %
23		Vehicle	AVE [13]	5394	75.7 m	9.5	77.96 %
24		Vehicle	SCAAT	5416	**79.9 m**	**10.3**	**79.75 %**

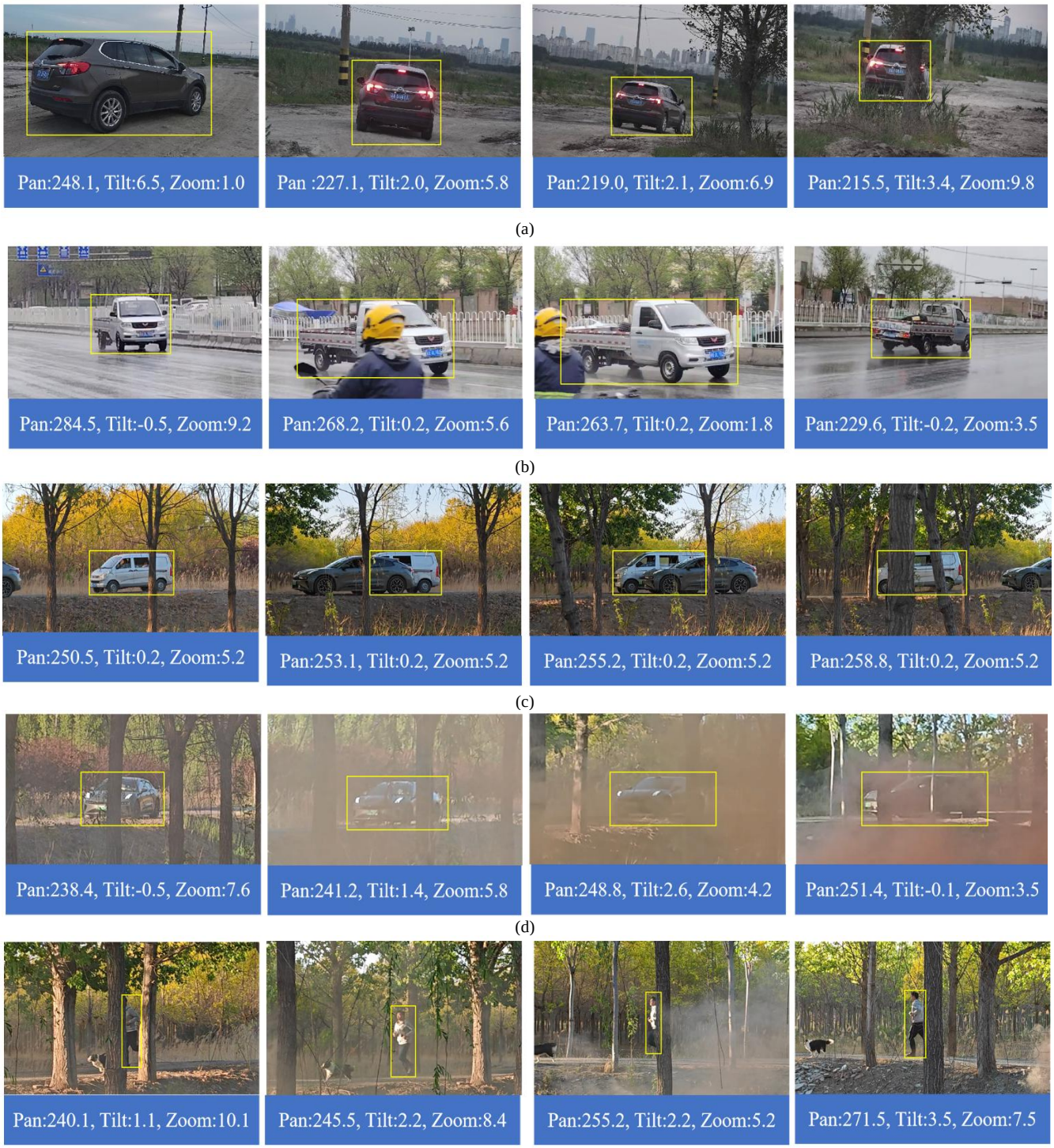
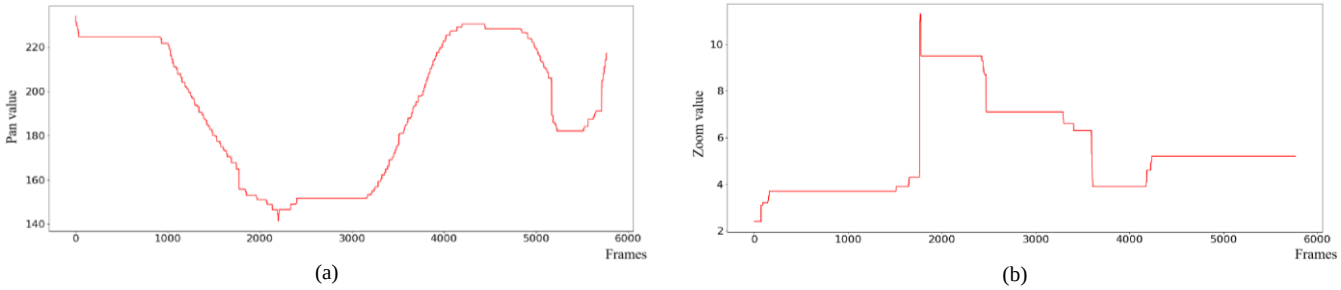
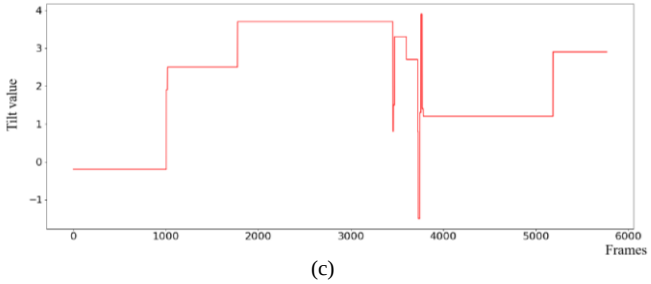


Fig. 12. Qualitative results of the proposed SCAAT tracker in real-world field tests: (a) Tracking the vehicle on a sunny day; (b) Tracking the vehicle under rain and occlusion; (c) Tracking the vehicle under occlusion; (d) Tracking the vehicle under smog and occlusion; (e) Tracking the person under smog and occlusion.





(c)

Fig. 13. Variational pans, tilts, and zooms with frames in experiments. A total of 5766 tracking frames captured: (a) Variational pans with frames; (b) Variational tilts with frames; (c) Variational zooms with frames.

V. CONCLUSIONS

In this work, a PTZ-based SCAAT method is proposed to track moving objects with variable speed variations. The object detection and tracking are developed and a PTZ camera controller is constructed in the spherical coordinate system. Start-time delay and start-angle delay are calibrated in the PTZ camera control to improve object tracking performance. Moreover, to adapt various objects' speeds and avoid the PTZ camera shaking under different zooms, an adaptive tracking window is proposed to keep the object within the camera FOV. An experimental evaluation has been performed and the results demonstrate high performance of the proposed PTZ-based SCAAT method with a good improvement in tracking accuracy and stability for moving objects with variable speeds. Spherical coordinates were utilised for active tracking while accounting for the gimbal camera startup delay. This approach can achieve an 8.98 % average increase in the effective distance, an 8.37 % improvement in the effective zoom, and a 1.62 % rise in the stability ratio, and a 2.4 % enhancement in the tracking precision.

FUTURE WORKS

The utilisation of the spherical coordinate system and the adaptive tracking window are the main factors for improving tracking performance. To this end, it is possible to apply the SCAAT method to different types of PTZ cameras. In the future, reinforcement learning will be applied to the proposed method for different types of PTZ cameras.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] L. Zhang, H. Han, M. Zhou, Y. Al-Turki, and A. Abusorrah, "An improved discriminative model prediction approach to real-time tracking of objects with camera as sensors", *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17308–17317, 2021. DOI: 10.1109/JSEN.2021.3079214.

[2] L. He, G. Liu, G. Tian, J Zhang, and J. Ze, "Efficient multi-view multi-target tracking using a distributed camera network", *IEEE Sensors Journal*, vol. 20, no. 4, pp. 2056–2063, 2020. DOI: 10.1109/JSEN.2019.2949385.

[3] H. Kim, D. Kim, and S.-M. Lee, "Marine object segmentation and tracking by learning marine radar images for autonomous surface vehicles", *IEEE Sensors Journal*, vol. 23, no. 9, pp. 10062–10070, 2023. DOI: 10.1109/JSEN.2023.3259471.

[4] L. He, X. Liao, W. Liu, X. Liu, P. Cheng, and T. Mei, "Fastreid: A Pytorch toolbox for general instance re-identification", in *Proc. of the 31st ACM International Conference on Multimedia*, 2023, pp. 9664–9667. DOI: 10.1145/3581783.3613460.

[5] C. Liu *et al*., "City-scale multi-camera vehicle tracking guided by crossroad zones", in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4124–4132. DOI: 10.1109/CVPRW53098.2021.00466.

[6] M. Wu, Y. Qian, C. Wang, and M. Yang, "A multi-camera vehicle tracking system based on city-scale vehicle re-ID and spatial-temporal information", in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4072–4081. DOI: 10.1109/CVPRW53098.2021.00460.

[7] Z. Hu *et al*., "All-day image alignment for PTZ surveillance based on correlated siamese neural network", *Signal, Image and Video Processing*, vol. 18, no. 1, pp. 615–624, 2024. DOI: 10.1007/s11760-023-02720-x.

[8] A. Dionigi, S. Felicioni, M. Leomanni, and G. Costante, "D-VAT: End-to-end visual active tracking for micro aerial vehicles", *IEEE Robotics and Automation Letters*, vol. 9 no. 6, pp. 5046–5053, 2024. DOI: 10.1109/LRA.2024.3385700.

[9] T. Zhang, Z. Li, Q.Wang, K. Shimasaki, I. Ishii, and A. Namiki, "DoF-extended zoomed-in monitoring system with high-frame-rate focus stacking and high-speed pan-tilt adjustment", *IEEE Sensors Journal*, vol. 24, no. 5, pp. 6765–6776, 2024. DOI: 10.1109/JSEN.2024.3351202.

[10] J. D. Muñoz, J. Ruiz-Santaquiteria, O. Deniz, and G, Bueno, "Weapon detection using PTZ cameras", in *Robotics, Computer Vision and Intelligent Systems. ROBOVIS 2024. Communications in Computer and Information Science*, vol. 2077. Springer, Cham, 2024, pp. 100–114. DOI: 10.1007/978-3-031-59057-3_7.

[11] F. Zhong, P. Sun, W. Luo, T. Yan, and Y. Wang, "AD-VAT+: An asymmetric dueling mechanism for learning and understanding visual active tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 5, pp. 1467–1482, 2021. DOI: 10.1109/TPAMI.2019.2952590.

[12] W. Luo *et al.*, "End-to-end active object tracking and its real-world deployment via reinforcement learning", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 6, pp. 1317–1332, 2019. DOI: 10.1109/TPAMI.2019.2899570.

[13] H. Fang, H. Liu, J. Wen, Z. Yang, J. Li, and Q. Han, "Automatic visual enhancement of PTZ camera based on reinforcement learning", *Neurocomputing*, vol. 626, art. 129531, 2025. DOI: 10.1016/j.neucom.2025.129531.

[14] P. D. Z. Varcheie and G. A. Bilodeau, "People tracking using a network-based PTZ camera", *Machine Vision and Applications*, vol. 22, no. 4, pp. 671–690, 2011. DOI: 10.1007/s00138-010-0300-1.

[15] K. Yun, H. Kim, K. Bae, and J. Park, "Unsupervised moving object detection through background models for PTZ camera", in *Proc. of 2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 3201–3208. DOI: 10.1109/ICPR48806.2021.9413085.

[16] A. Zhu, J. Yang, and W. Yu, "A novel target tracking method of unmanned drones by gaze prediction combined with YOLO algorithm", in *Proc. of 2021 IEEE International Conference on Unmanned Systems (ICUS)*, 2021, pp. 792–797. DOI: 10.1109/ICUS52573.2021.9641499.

[17] J. Hu, C. Zhang, S. Xu, and C. Chen, "An invasive target detection and localization strategy using pan-tilt-zoom cameras for security applications", in *Proc. of 2021 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, 2021, pp. 1236–1241. DOI: 10.1109/RCAR52367.2021.9517521.

[18] S. Fang and H. Li, "Multi-vehicle cooperative simultaneous LiDAR SLAM and object tracking in dynamic environments", *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 9, pp. 11411–11421, 2024. DOI: 10.1109/TITS.2024.3360259.

[19] P. Nguyen, K. G. Quach, C. N. Duong, S. L. Phung, N. Le, and K. Luu, "Multi-camera multi-object tracking on the move via single-stage global association approach", *Pattern Recognition*, vol. 152, art. 110457, 2024. DOI: 10.1016/j.patcog.2024.110457.

[20] G.-J. Lee, S.-W. Jang, and G.-Y. Kim, "Pupil detection and gaze tracking using a deformable template", *Multimedia Tools and Applications*, vol. 79, pp. 12939–12958, 2020. DOI: 10.1007/s11042-020-08638-7.

[21] C. Qiu, Z. Wu, S. Kong, and J. Yu, "An underwater micro cable-driven pan-tilt binocular vision system with spherical refraction calibration", *IEEE Transactions on Instrumentation and Measurement*, vol. 70, art. no. 5010813, pp. 1–13, 2021. DOI: 10.1109/TIM.2021.3082258.

[22] A. Tsoukalas, D. Xing, N. Evangeliou, N. Giakoumidis, and A. Tzes, "Deep learning assisted visual tracking of evader-UAV", in *Proc. of 2021 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2021, pp. 252–257. DOI: 10.1109/ICUAS51884.2021.9476720.

[23] Y. Li, L. Pan, and T. Cheng, "A camera PTZ control algorithm for

autonomous mobile inspection robot", in *Proc. of 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 2021, pp. 962–967. DOI: 10.1109/ICBAIE52039.2021.9389970.

[24] J. Ji, "PTZ camera control based on dynamic object detection", M.S. thesis, North China Electric Power University, Beijing, China, 2021. DOI: 10.27140/d.cnki.ghbbu.2021.000528.

[25] D. Li, W. Li, S. Chen, and S. Zhao, "Intelligent robot for early childhood education", in *Proc. of 2020 8th International Conference on Information and Education Technology*, 2020, pp. 142–146. DOI: 10.1145/3395245.3396420.

[26] S. Hu, K. Shimasaki, M. Jiang, T. Senoo, and I. Ishii, "A simultaneous multi-object zooming system using an ultrafast pan-tilt camera", *IEEE Sensors Journal*, vol. 21, no. 7, pp. 9436–9448, 2021. DOI: 10.1109/JSEN.2021.3054425.

[27] L. Shen, S. Hu, K. Shimasaki, T. Senoo, and I. Ishii, "Simultaneous multi-face zoom tracking for 3-D people-flow analysis with face identification", in *Proc. of 2020 16th International Conference on Mobility, Sensing and Networking (MSN)*, 2020, pp. 410–417. DOI: 10.1109/MSN50589.2020.00073.

[28] A. Laucka, D. Andriukaitis, "Research of the Defects in Anesthetic Masks", *Radioengineering*, vol. 24, no. 4, pp. 1033–1043, 2015. DOI: 10.13164/re.2015.1033.

[29] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN", in *Proc. of 2017 IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988. DOI: 10.1109/ICCV.2017.322.

[30] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017. DOI: 10.1109/TPAMI.2016.2577031.

[31] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection", in *Proc. of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988. DOI: 10.1109/ICCV.2017.324.

[32] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection", in *Proc. of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91.

[33] M. Durve *et al.*, "DropTrack—Automatic droplet tracking with YOLOv5 and DeepSORT for microfluidic applications", *Physics of Fluids*, vol. 34, no. 8, p. 082003, 2022. DOI: 10.1063/5.0097597.

[34] B. Yang and R. Nevatia, "Online learned discriminative part-based appearance models for multi-human tracking", in *Computer Vision – ECCV 2012. ECCV 2012. Lecture Notes in Computer Science*, vol. 7572. Springer, Berlin, Heidelberg, 2012, pp. 484–498. DOI: 10.1007/978-3-642-33718-5_35.

[35] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T.-K. Kim, "Multiple object tracking: A literature review", *Artificial Intelligence*, vol. 293, art. 103448, 2021. DOI: 10.1016/j.artint.2020.103448.