

Tracking Control of a Pan–Tilt Camera System in Maritime Search and Rescue

Nguyen Duc Hai Duong^{*} , Do Trung Kien , Bui Hai Dang , Pham Quang Ninh , and Nguyen Van Hung 

Abstract: Maritime search and rescue requires target detection and tracking systems that are highly accurate and capable of achieving the shortest possible response time. Therefore, pan–tilt–zoom (PTZ) camera systems must satisfy these requirements. This study proposes an intelligent pan–tilt camera system for maritime search and rescue under conditions of fluctuating water surfaces, changing illumination, and targets that are easily obscured by waves. Previous methods mainly relied on PTZ tracking, You Only Look Once version 8 (YOLOv8) combined with tracking algorithms, and several image enhancement techniques; however, they remained limited in terms of real–time stability and adaptability to nonlinear noise in the marine environment. In this paper, YOLOv8 is integrated for human detection, while the ByteTrack algorithm is used to maintain object identity and estimate the image–centering error. An optimized Adaptive Neuro–Fuzzy Inference System (ANFIS) controller is designed for the two pan–tilt axes based on red–green–blue (RGB) and thermal imaging. Experimental results in four scenarios demonstrate that the target tracking system is stable, with the error gradually decreasing to acceptable limits. The YOLOv8 model achieved a precision above 0.8, a recall of approximately 0.7–0.75, and a mean Average Precision at an Intersection over Union threshold of 0.50 (mAP@50) of approximately 0.7. In real–world testing, the pan error was reduced from –674 pixels to 5 pixels.

Keywords: ANFIS, ByteTrack, maritime search and rescue, pan–tilt camera, target tracking, YOLOv8.

1. INTRODUCTION

Automatic surface surveillance systems installed on ships and at seaports have attracted increasing attention, particularly for maritime search and rescue (SAR) and safety monitoring. The rapid growth of maritime transportation and the increasing diversity of vessels have made maritime operations more complex and potentially hazardous [1, 2]. In addition, adverse weather and sea conditions can considerably reduce the effectiveness of conventional visual surveillance. Variations in illumination, water–surface reflections, image blur, sea spray, and wave–induced target occlusion are among the major factors that make the detection and continuous tracking of small targets, particularly humans in water, challenging. Therefore, robust automatic surveillance systems capable of maintaining reliable detection and tracking under uncertain maritime conditions are required. [3, 4]. Therefore, the task is to develop automatic surveillance systems with high adaptability and stable operation in harsh environmental conditions.

Recent studies have increasingly investigated multisensor solutions combining RGB cameras, thermal imaging, and LiDAR for maritime perception and SAR applications [5]. RGB cameras provide detailed visual information under favorable illumination, thermal cameras improve target visibility at night or under low–light conditions, while LiDAR can provide complementary geometric and distance information. Such sensing systems have also been integrated into USVs, ASVs, and other autonomous platforms to detect small floating objects or persons in water. However, reliable maritime surveillance requires not only target detection but also continuous target tracking and accurate pan–tilt control when the target moves or is temporarily obscured by waves. [6, 7].

Deep–learning–based object detection and tracking have significantly improved the performance of vision–based surveillance systems. In work [8], Feng Lu et al. developed a PTZ algorithm for variable velocity object tracking, focusing on tracking stability and object centering using PTZ control combined with image processing to predict target position. In research [9],

Manuscript received March 14, 2026; received in revised form July 14, 2026; accepted August 15, 2026; available online September 30, 2026.

Nguyen Duc Hai Duong, Do Trung Kien, Bui Hai Dang, and Pham Quang Ninh are the students of the Faculty of Electrical and Electronic Engineering, Vietnam Maritime University, Haiphong, Vietnam (e–mails: duong106976@st.vimaru.edu.vn, kien108459@st.vimaru.edu.vn, dang107133@st.vimaru.edu.vn, and ninh109695@st.vimaru.edu.vn). Nguyen Van Hung is with the Faculty of Electrical and Electronic Engineering, Vietnam Maritime University, Haiphong, Vietnam (e–mail: hung.ddt@vimaru.edu.vn).

^{*} Corresponding author.

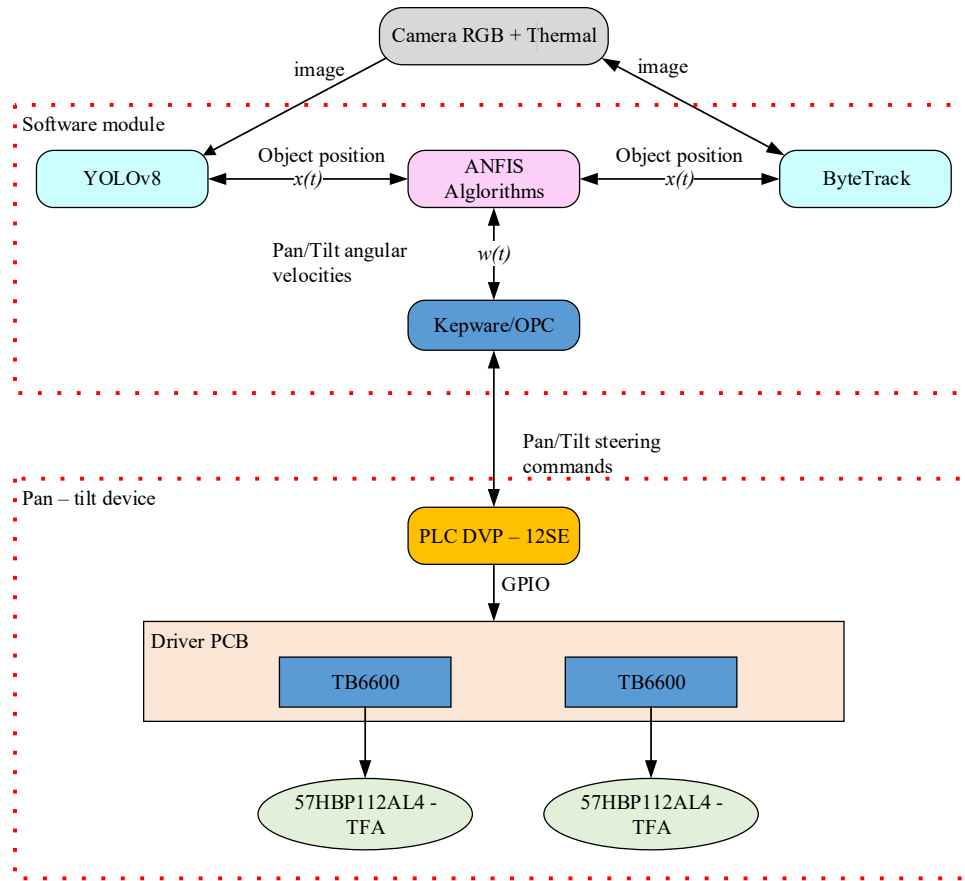


Fig. 1. System structure.

Ming Lui et al. used knowledge distillation and data augmentation to train BicycleGAN for image enhancement, developing adaptive YOLO (Ultralytics YOLO family) in combination with trackers BoT–SORT and ByteTrack to support PTZ imaging and tracking. Deep Learning models have provided good support for cameras to detect objects [9]–[11].

Stabilizing and tracking objects for cameras requires controllers that are adaptable and responsive to real-time control requirements. ANFIS provides an effective solution by combining the advantages of neural networks and fuzzy logic, allowing the system to learn from data and make flexible control decisions [12]. ANFIS in pan and tilt camera control can adapt to fluctuating marine environmental conditions by combining neural networks and fuzzy logic, enabling adaptive and flexible control decisions [13]–[15].

Several studies have proposed a new approach through the integration of computer vision and intelligent control systems to improve efficiency and response speed in maritime emergencies. The YOLO model combines computation, determining the distance from the person's position to the center of the image as input for the ANFIS controller. The system is evaluated under static and dynamic target motion, temporary target loss in a swimming-pool environment, and real coastal conditions involving wave-induced occlusion and varying illumination. The objective is to investigate both detection and closed-loop tracking performance under progressively more realistic maritime conditions. [14],

[18].

New contributions of the research:

(i) Firstly, it proposes an optimized Adaptive Neuro-Fuzzy Inference System (ANFIS)-based controller for the pan-tilt mechanism, which improves robustness against nonlinear disturbances in the marine environment.

(ii) Secondly, it develops an integrated vision-based tracking framework that combines YOLOv8 for human detection and ByteTrack for target association, while also estimating image-centering error and target distance for control purposes.

(iii) Thirdly, the proposed system is validated through four experimental scenarios, including laboratory, swimming-pool, and real-world coastal tests, with the additional aim of future deployment on the Vietnam–Korea VMU vessel to further demonstrate its stability and accuracy in comparison with conventional methods.

The remainder of this paper is organized as follows: Section 2 presents the proposed system architecture. Section 3 describes the object detection and tracking model and the theoretical basis for distance calculation. Section 3 describes the experimental scenarios and the response of the ANFIS controller. Finally, Section 4 concludes the paper and discusses future development directions.

2. TARGET IDENTIFICATION AND COORDINATE DETERMINATION

This section is based on the YOLOv8 model, which

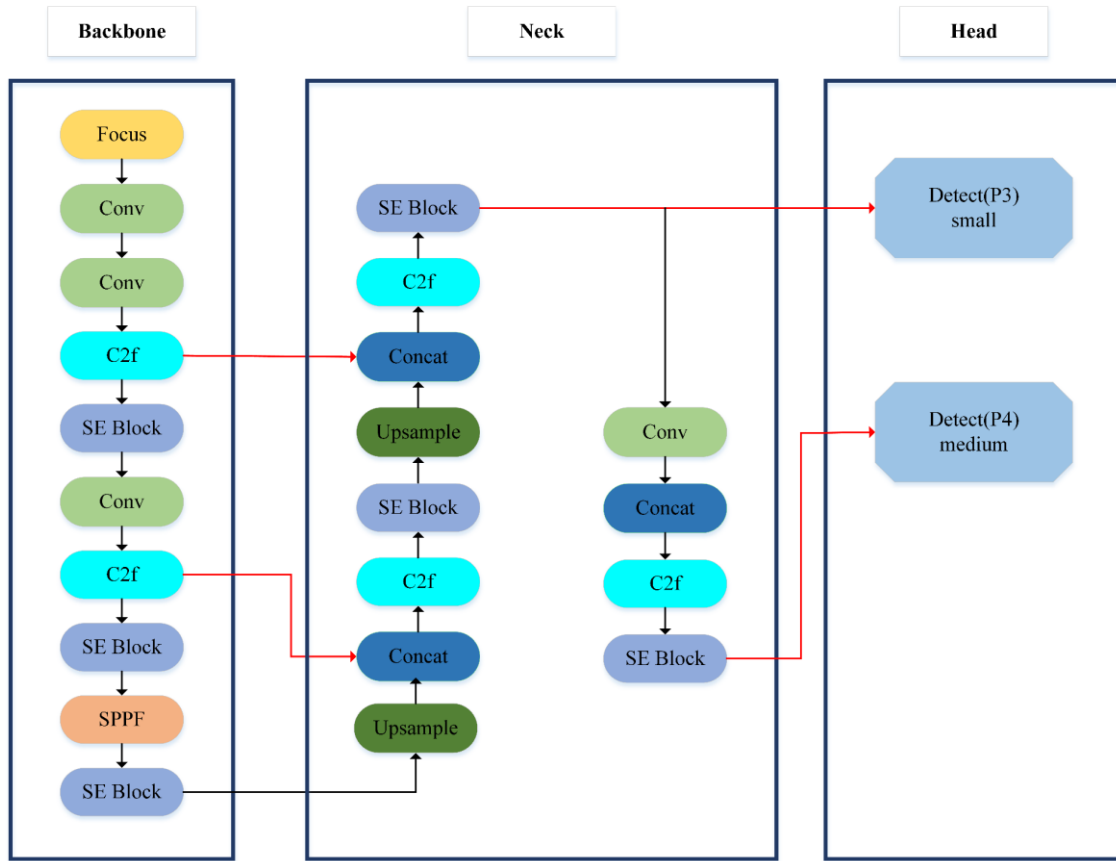


Fig. 2. Architecture of the proposed YOLOv8-SE model for maritime human-target detection.

allows the system to detect, recognize, and determine the coordinates of objects on a 2D plane, thereby calculating parameters as input to the neural network model. Section 2.1 presents the overall architecture of the system. Section 2.2 describes the structure of the YOLOv8 model for object recognition. Section 2.3 addresses the object tracking method and normalization of center coordinates.

2.1. Propose a system structure

Fig. 1 illustrates the overall structure of the system that allows for real-time tracking and control of the motors of the two object-following axes. The system consists of three main parts:

(i) Input block: consists of two types of cameras that allow observation of objects both day and night to provide color and thermal imaging data.

(ii) Software block: Includes the YOLOv8 model for object recognition and framing, which the ByteTrack algorithm uses to track and calculate the coordinate misalignment between the camera center and the object's image center. This misalignment is then used as input to the ANFIS model to send control signals to Kepware via Modbus TCP/IP protocol.

(iii) The actuator block includes a PLC, driver, and stepper motors to control the two pan and tilt axes for object tracking.

After constructing the structural diagram, the authors proceeded to build the YOLOv8 model structure for object detection and recognition.

2.2. YOLOv8 model

Fig. 2 illustrates the structure of the YOLOv8 model used in the study, along with its stability and reliability. Several studies have addressed the process of detecting people on the water surface based on image processing algorithms and deep learning models. The limitations of previous methods, which primarily used RGB camera images for motion and shape detection, are significant. However, operating in harsh environments with poor lighting, fog, and large waves can cause image noise. Therefore, several image processing methods such as UWCNN, U-net, and DnCNN have been developed to enhance image quality and reduce noise [19]. In the previous study, the authors used the UnCNN method to reduce noise in the underwater human detection model. To improve the representation of human-target features under complex maritime backgrounds, an SE attention block is introduced into the YOLOv8 architecture, as shown in Fig. 2. The SE module adaptively recalibrates channel-wise feature responses through squeeze and excitation operations, thereby assigning larger weights to informative feature channels and suppressing less relevant background information. This modification is intended to improve the robustness of human-target detection under illumination variation, water-surface reflections, and partial target occlusion.

2.3. Algorithm

YOLOv8 is capable of highly accurate object detection based on independent and unlinked sampling

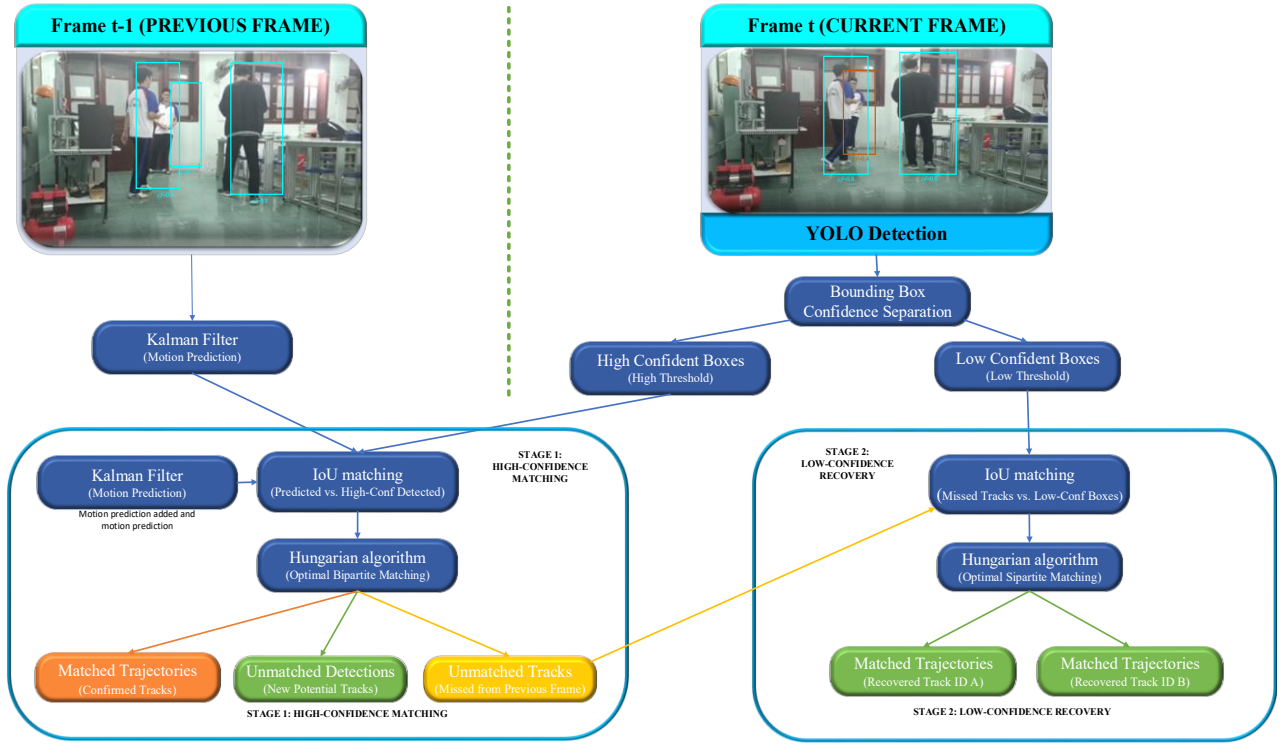


Fig. 3. Framework of multi-object tracking with high- and low-confidence associatio.

cycles, relying on data from previous frames to predict the object's trajectory. Therefore, in many cases, it loses position and deviates from the target. To overcome this limitation, the authors propose the ByteTrack algorithm, which offers high performance for fast movements and is implemented on embedded devices as shown in Fig. 3. The algorithm's principle, based on the Kalman filter for predicting the bounding box, currently includes two stages as follows [20], [21]:

–Phase 1: ByteTrack will match the predicted bounding boxes with YOLO labels and check the match rate based on the IoU index [21], [22].

–Phase 2: If matching is not possible, the algorithm will match with bounding boxes of lower confidence.

Through these two stages, once an object is detected, it will be assigned an ID for tracking and calculating control parameters for ANFIS.

Determine the object's coordinates based on the camera frame as shown in Fig. 4, where the width W_{frame} , height H_{frame} and frame coordinates (c_x, c_y) are determined as follows:

$$c_{x0} = \frac{W_{frame}}{2}; \quad c_{y0} = \frac{H_{frame}}{2} \quad (1)$$

YOLO sends back the range parameters as a matrix $[x_{max}, y_{max}, x_{min}, y_{min}]$. Coordinates of the center of the boundary frame (c_x, c_y) or the location of the person extracted from the boundary frame:

$$c_x = \frac{x_{max} + x_{min}}{2}; \quad c_y = \frac{y_{max} + y_{min}}{2} \quad (2)$$

The distance error, measured in pixels, of each pan and tilt axis in the pixel coordinate system is determined

by the distance from the center of the frame to the center of the boundary frame.

$$\begin{aligned} e_{pan} &= c_{x0} - c_x \\ e_{tilt} &= c_{y0} - c_y \end{aligned} \quad (3)$$

Simultaneously, the size of the boundary frame is calculated to help the system estimate the relative distance between the camera and the target in 3D space:

$$scale = \frac{(x_{max} - x_{min})(y_{max} - y_{min})}{W_{frame}H_{frame}} \quad (4)$$

The normalization process for pan and tilt axis misalignment is based on the clipping value limit function shown in Fig. 5.

This function limits the value within an arbitrary range based on the min and max values, acting as a saturation element. Therefore, the normalization process for the positional error of the two pan and tilt axes is calculated as follows [23]–[25]:

$$e_{norm} = \begin{cases} 1 & \text{if } \frac{e_{raw}}{e_{max}} \geq 1 \\ \frac{e_{raw}}{e_{max}} & \text{if } -1 < \frac{e_{raw}}{e_{max}} < 1 \\ -1 & \text{if } \frac{e_{raw}}{e_{max}} \leq -1 \end{cases} \quad (5)$$

where, e_{norm} is a normalized coordinate error, e_{raw} The pan and tilt axis misalignment needs to be calibrated, e_{max} is the maximum value of e_{raw} in the pixel system.

In cases where the object has been assigned a bounding box and has a confidence score, it will be normalized using the Linear Mapping method:

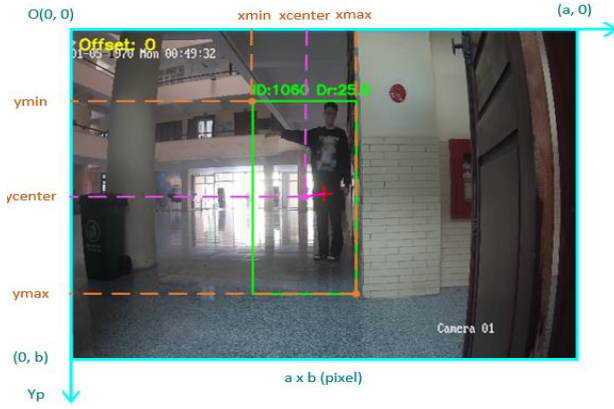


Fig. 4. Camera coordinate system for human detection.

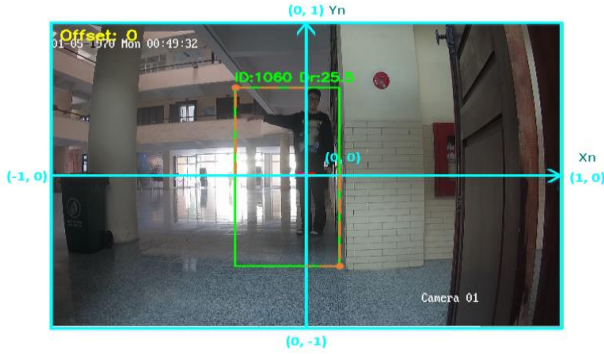
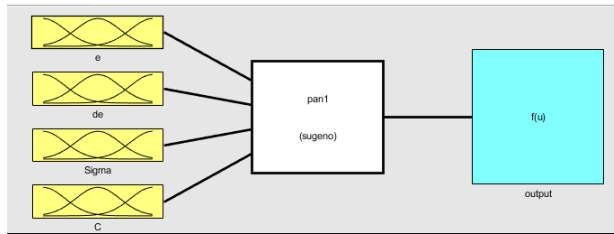
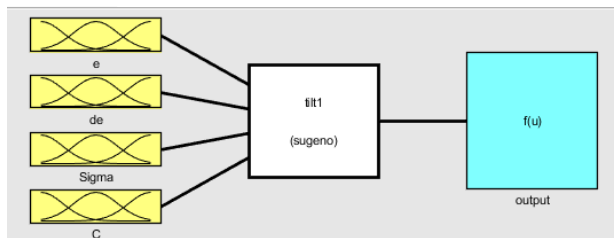


Fig. 5. Camera coordinate system for human detection.



(a) Structure of the PAN axis.



(b) Structure of the TILT axis.

Fig. 6. ANFIS controller.

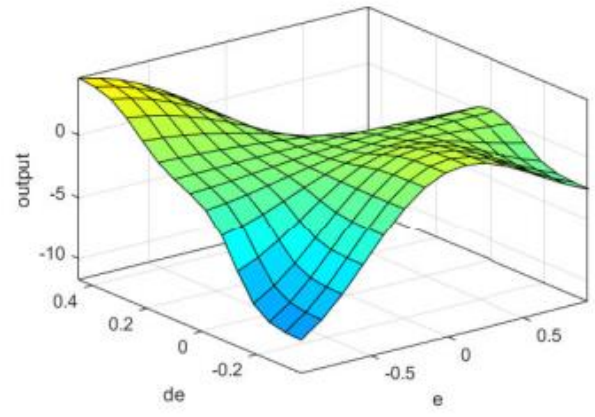
$$s_{norm} = 2scale - 1 \quad (6)$$

$$cf_{norm} = 2cf - 1 \quad (7)$$

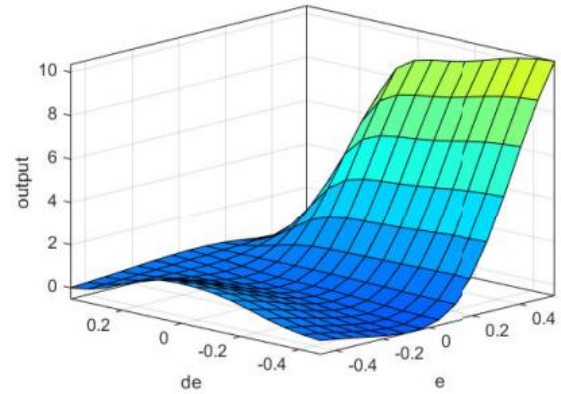
where, cf_{norm} is a standardized confidence point, cf is the initial reliability.

The rate of change of positional error will be calculated based on the normalized and smoothed coordinate deviations [20]–[22], [26]:

$$de = en_t - en_{t-1} \quad (8)$$



(a) Structure of the PAN axis.



(b) Structure of the TILT axis.

Fig. 7. ANFIS surface of rules.

where, en_t is the normalized distance error that has been smoothed at time t , en_{t-1} is the smoothed normalized distance error at time $t-1$.

In practice, in tilting problems, bounding boxes can encounter vibration noise, causing them to oscillate and affecting the accuracy of the algorithm, specifically the position coordinates of the tracked object, resulting in motor jerks. In this paper, an Exponential Moving Average (EMA) filter is proposed as a first-order low-pass filter to smooth out coordinate deviations [27]:

$$EMA_t = \alpha \cdot P_t + (1 - \alpha) \cdot EMA_{t-1} \quad (9)$$

where, EMA_t denotes the smoothed value at the current time t , EMA_{t-1} represents the smoothed value from the preceding period, and P_t is the raw value to be smoothed. The parameter $\alpha \in [0, 1]$ serves as the smoothing coefficient, which is calibrated to optimize the balance between smoothing efficacy and the system's lag response.

3. TRACKING CONTROL USING ANFIS FUZZY NEURAL NETWORK

In this section, the research group applies the ANFIS fuzzy neural network method to control and track the target. The dataset is collected in real time and

Table 1. Experimental scenarios for performance evaluation.

Scenario	Title	Objectives	Operational conditions
1	Static target tracking	Evaluate convergence and steady-state accuracy.	A stationary target is initialized with a 100–150 pixel offset.
2	Dynamic target tracking	Evaluate tracking under continuous target motion.	The subject moves along nonlinear paths at varying speeds.
3	Swimming-pool target-loss recovery	Evaluate target re-detection and tracking recovery after temporary target loss.	The subject moves in a swimming pool and is intermittently obscured by partial submergence, water-surface disturbances, or controlled occlusion.
4	Real-world coastal tracking	Evaluate robustness under realistic maritime disturbances.	Beach testing with waves, low light, image blur, and intermittent target occlusion.



Fig. 8. Test hardware model for target tracking camera.

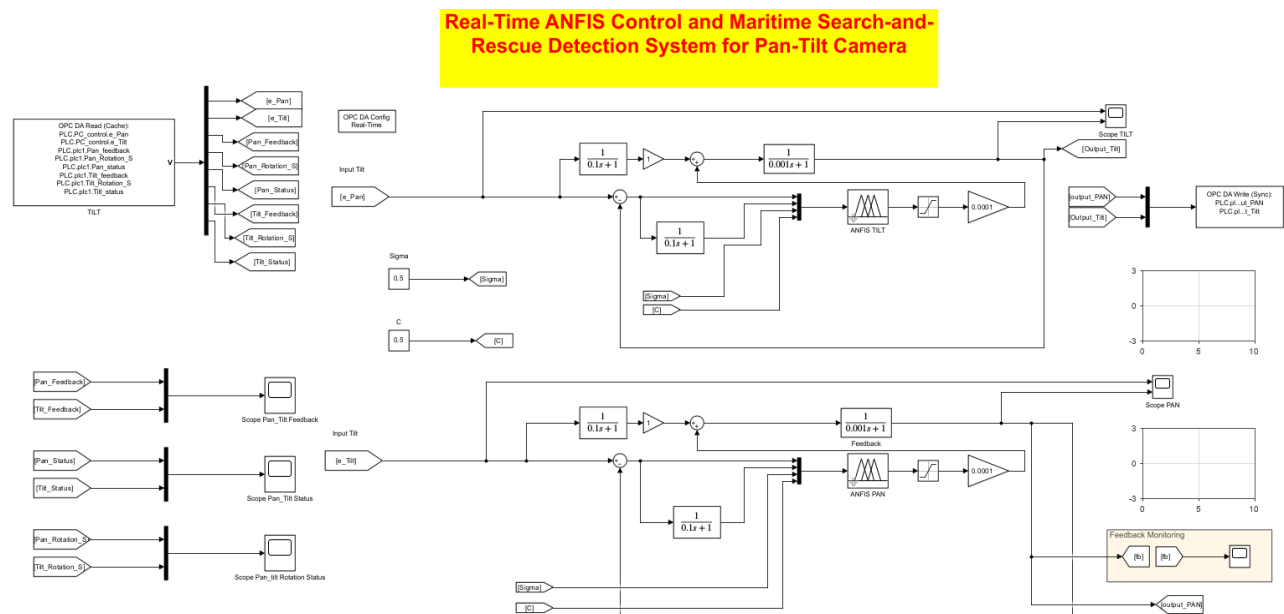


Fig. 9. ANFIS real-time control interface for pan-tilt camera systems in maritime search and rescue applications.

normalized to the range $[-1, 1]$ and then fed as input to the ANFIS model for training. Section 3.1 presents the data collection process for training the controller, section

3.2 presents the ANFIS controller structure of the model, and section 3.3 builds the rules for the ANFIS model. Finally, the results after training the ANFIS model are

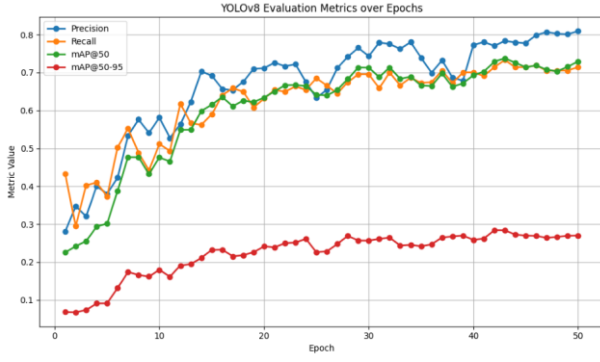


Fig. 10. YOLOv8 model performance.

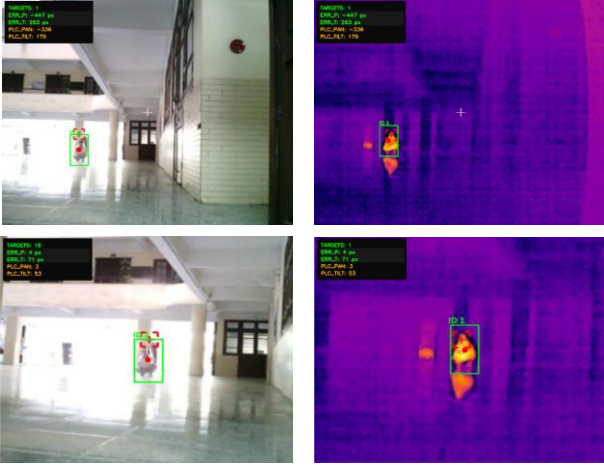
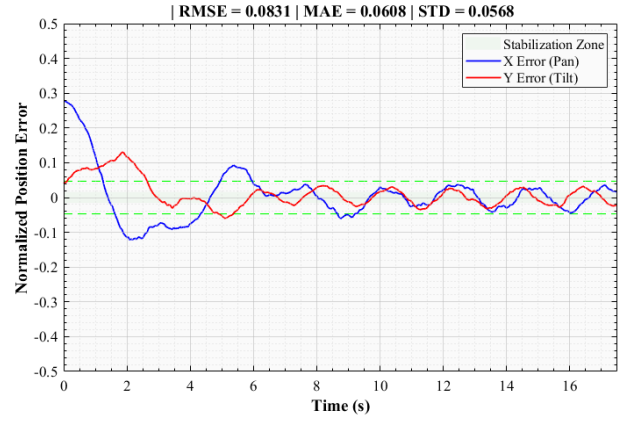
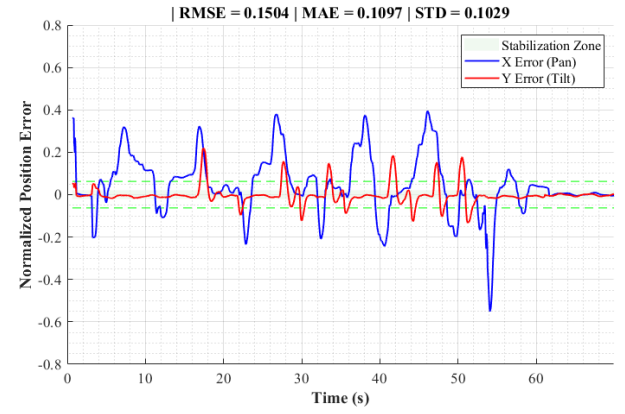


Fig. 11. RGB and thermal target detection – Scenario 1.



(a) Simulation.



(b) Experiment.

Fig. 12. Results of scenario 1.

presented.

3.1. Training data

The ANFIS controller will perform well and flexibly when the training data needs to be accurate and diverse, encompassing many different scenarios. Therefore, the authors developed a two-axis PAN-TILT stepper motor model that tracks objects, controlled by a fine-tuned PID controller using the Ziegler-Nichols method, and implemented various behaviors to generate a dataset with smooth and stable object-center tracking within an acceptable error range. The collected data was then fed into the ANFIS training process. The Ziegler-Nichols method aims to optimize the PID controller based on critical parameters: K_u is ultimate gain and T_u is ultimate period [28, 29].

In real-time processing models, the constraint between the speed of the recognition algorithm and the response speed of the controller and motor must be compatible based on the following requirements:

$$T_{inf} \ll \frac{1}{f_s} \quad (10)$$

where, T_{inf} denotes the YOLO inference latency; f_s represents the control sampling frequency.

In addition, the model also considers the electromechanical properties of the system for calculation and control. The 57HBP112AL4-TFA stepper motor is a 2-phase type, so the electrical model

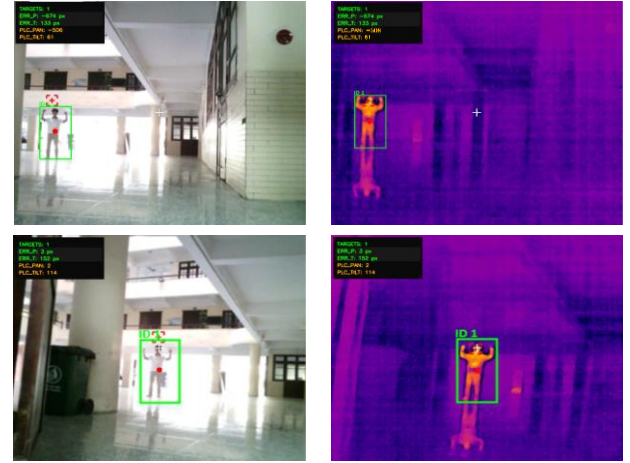


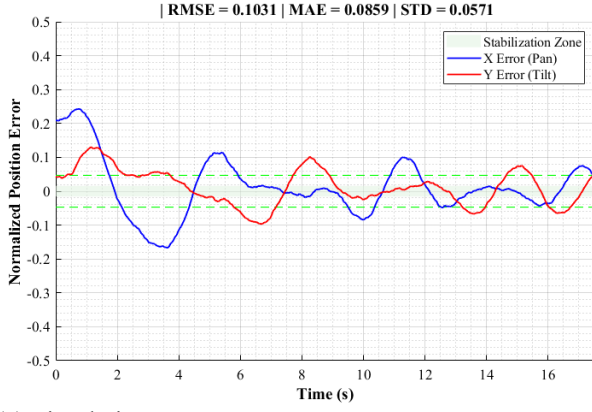
Fig. 13. Visual and thermal target detection and tracking results.

is described by the following system of equations [22]:

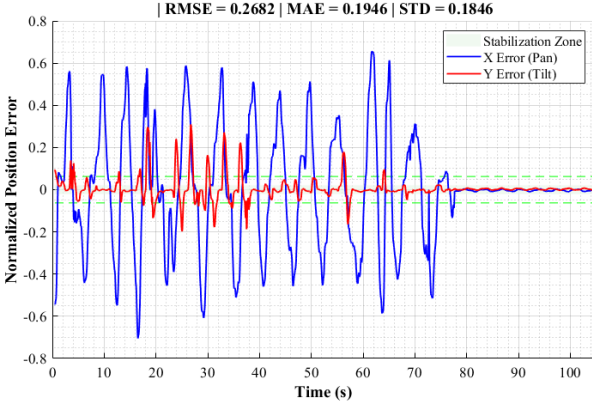
$$U_a(t) = R_a i_a(t) + L_a \frac{di_a(t)}{dt} - K_b \omega(t) \sin(N_T \theta(t)) \quad (11)$$

$$U_b(t) = R_a i_b(t) + L_a \frac{di_b(t)}{dt} - K_b \omega(t) \sin(N_T \theta(t)) \quad (12)$$

where, U_a, U_b are the voltages applied to phases A and B; R_a, L_a denote the resistance and inductance of the stator windings; i_a, i_b represent the phase currents;

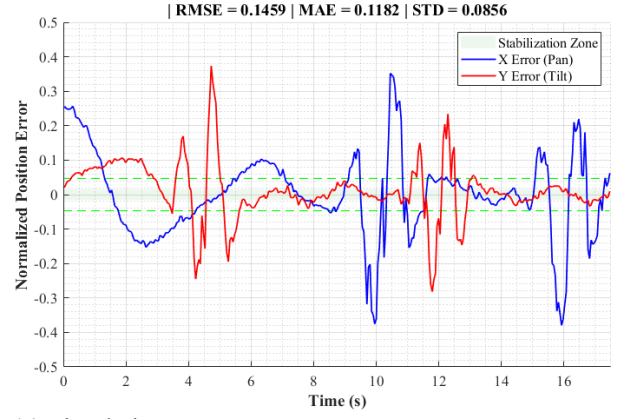


(a) Simulation.

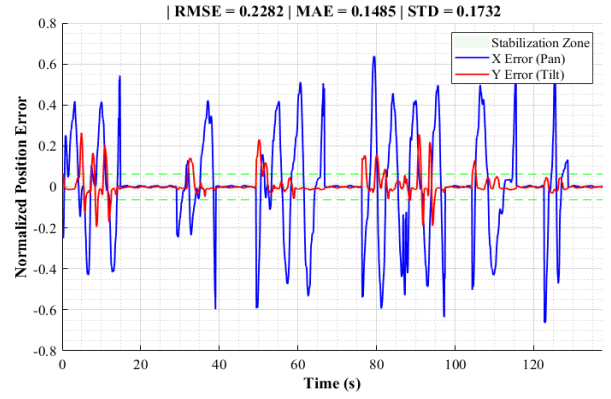


(b) Experiment.

Fig. 14. Results of scenario 2.



(a) Simulation.



(b) Experiment.

Fig. 16. Results of scenario 3.

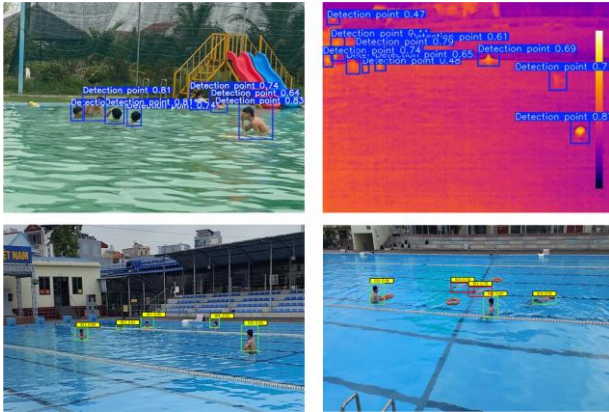


Fig. 15. RGB and thermal target detection – Scenario 3.



Fig. 17. RGB and thermal target detection – Scenario 4.

$e_b = K_b \omega$ is the back-electromotive force, and $N_r = 50$ indicates the number of rotor teeth.

In addition, the engine's mechanical model is expressed through the overall torque of the system based on the dynamic equations:

$$\begin{aligned} \tau_m(t) &= J \frac{d\omega(t)}{dt} + B\omega(t) + \tau_L(t) \\ &= (J + J_L) \frac{d\omega(t)}{dt} + B\omega(t) \end{aligned} \quad (13)$$

where, $\tau_m(t) = K_m [-i_a(t) \sin(N_r \theta(t)) + i_b(t) \cos(N_r \theta(t))]$ denotes the electromagnetic torque generated by the

stepper motor; J is the rotor moment of inertia; J_L represents the load moment of inertia; B is the viscous friction coefficient; and τ_L signifies the external load torque, determined by the specific camera model mounted on the shaft.

3.2. ANFIS controller

The ANFIS controller combines the reasoning capabilities of fuzzy logic with the self-learning and optimization capabilities of artificial neural networks.

Table 2. Summary of tracking performance under different evaluations.

Scenario	Test condition	RMSE	MAE	STD	Main result
1	Static target tracking	0.1504	0.1097	0.1029	Stable convergence
2	Dynamic target tracking	0.2682	0.1946	0.1846	Continuous tracking under target motion
3	Swimming-pool target-loss recovery	0.2282	0.1485	0.1732	Target re-detection and tracking recovery
4	Real-world coastal tracking	0.1282	0.0950	0.0862	Tracking under waves and intermittent occlusion

Table 3. Comparison with representative maritime detection and tracking approaches.

Method	Main task	Main quantitative result	Occlusion/real maritime validation
Chen et al. [36]	Maritime target tracking	RMSE, RMSPE, MAD, MAPE < 10	Yes/Yes
Ponzini et al. (2025) [37]	Human detection in marine SAR	Precision = 93.5%, Recall = 94.2%	Yes/Yes
MES-YOLO (2025) [38]	Maritime SAR detection	mAP@50 = 87.1%	Maritime dataset
HPRT-YOLO (2026) [39]	Lightweight maritime SAR detection	P = 88.5%, R = 81.1%, mAP@50 = 85.6%, 94.9 FPS	Maritime dataset
JDR-CSTrack (2025) [40]	Maritime multi-object tracking	HOTA = 70.1, MOTA = 77.3, IDF1 = 85.7	Yes/Maritime videos
Proposed system	Human detection + pan-tilt tracking	P ≈ 81%, R ≈ 71%, mAP@50 ≈ 73%; RMSE = 0.1282	Yes/Yes

The ANFIS architecture is designed with rule-based inputs and outputs, ensuring that the number of rules does not exceed the number of output member functions. In this study, the ANFIS controller is configured with four inputs: determining the direction and velocity to be compensated (e), predicting the motion trend to reduce oscillations (de), estimating distance and self-tuning (σ), and evaluating the reliability of the feedback data (C). The system uses a first-order Sugeno model, where each fuzzy rule is described as follows [30]–[32]:

Rule 1: If e is A_1 , de is B_1 , σ is C_1 , and C is D_1 then $f_1 = p_1e + q_1de + r_1\sigma + s_1C + t_1$

Rule 2: If e is A_2 , de is B_2 , σ is C_2 , and C is D_2 then $f_2 = p_2e + q_2de + r_2\sigma + s_2C + t_2$

Rule 3: If e is A_2 , de is B_2 , σ is C_2 , and C is D_2 then $f_3 = p_3e + q_3de + r_3\sigma + s_3C + t_3$

Fig. 6 (a) and Fig. 6(b) shows the structure of the ANFIS controller with 5 layers [33].

Layer 1 is the blurring layer. Each node of Layer 1 produces members of the language label in the form of a generalized bell function [34], [35].

$$\text{bell}(x; a, b, c) = \frac{1}{1 + |(x - c) / a|^{2b}} \quad (14)$$

where, $\{a, b, c\}$ is the set of antecedent parameters. The shape of the bell function changes as the parameter values change.

$$O^2 = w_i = \mu_{a_i}(e)\mu_{b_i}(de)\mu_{c_i}(\sigma)\mu_{d_i}(C) \quad (15)$$

The product $\mu_{a_i}(e)\mu_{b_i}(de)\mu_{c_i}(\sigma)\mu_{d_i}(C)$ represents

the strength of a rule. e, de, σ and C in represent inputs with value a_i, b_i, c_i , and d_i , respectively

Layer 3 normalizes the activation intensity of all rules by calculating the usage ratio.

$$O^3 = \bar{w}_i = \frac{w_i}{\sum w_i} \quad \text{with } i = 1, 2 \quad (16)$$

The nodes at layer 4 are called adaptive nodes. These nodes compute the parametric function.

$$O^4 = \bar{w}_i f_i = (\bar{w}_i (p_i x_1 + q_i x_2 + r_i x_3 + s_i x_3 + t_i)) \quad (17)$$

where, p_i, q_i, r_i, s_i , and t_i are consequent parameters set. The parameters in this layer are referred to as consequent parameters.

Layer 5 is the output grade and is a synthesis of all the outputs from layer 4.

$$O^5 = \sum \bar{w}_i f_i = \frac{\sum w_i f_i}{\sum w_i} \quad \text{with } i = 1, 2 \quad (18)$$

The ANFIS model training data is generated from the results of a fuzzy logic controller based on expert rules. This data is smoothed through speed limiting and the removal of high-frequency oscillations to produce a smooth control characteristic curve.

3.3. Fuzzy rules and neral network

In this study, the authors consider the ANFIS controller as a normalized baseline controller. The system using ANFIS is designed to control the target-tracking object as shown in Fig. 7 (a) and Fig. 7(b). Training data is collected from a real-world operating scenario. All data has been pre-processed to remove

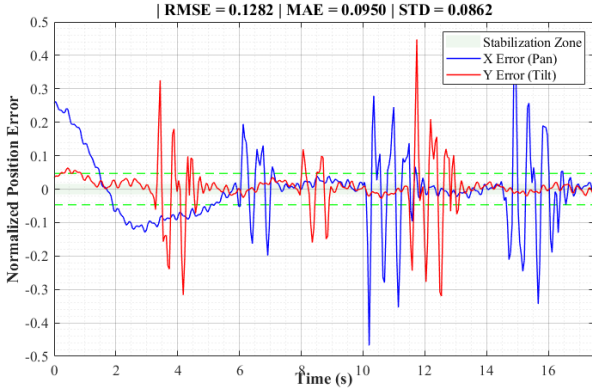


Fig. 18. Real-world coastal experimental results under wave-induced target occlusion.

noise and normalized to the range $[-1, 1]$. This normalization ensures compatibility between input and output variables and optimizes the control surface of the target-tracking system.

4. RESULTS AND DISCUSSION

In this section, the authors conduct controller testing with a real-world model. First, section 4.1 presents the construction of the test scenario for the ANFIS controller. The identification results in YOLO and the signal response of the PAN and TILT axes are shown in section 4.2. Finally, section 4.3 evaluates the model's performance and suggests directions for future development.

4.1. Experiment scenarios

The authors developed four test scenarios for the ANFIS model and the camera tracking system. Table 1 shows the content and evaluation methods for each scenario [21], [30], [34]. In addition, to evaluate the model, the authors' team used the following quantitative indicators:

Root mean square error (RMSE) is defined as

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (e_i)^2} \quad (19)$$

Mean absolute error (MAE) is given by

$$MAE = \frac{1}{N} \sum_{i=1}^N |e_i| \quad (20)$$

Standard deviation σ is utilized to demonstrate the stability of the error distribution around the zero-mean region.

Fig. 8 depicts the experimental laboratory hardware, including a pan-tilt camera model and a real-time ANFIS control interface for detecting and searching for people in the test area.

Fig. 9 shows the control interface built on Simulink/Matlab with functional blocks such as the data acquisition block from the PLC via the OPC protocol. In addition, control signal data from YOLOv8 calculates the image coordinate error and is sent via Simulink. Notably, the controller uses a reliability coefficient extracted directly from YOLO in Python to enhance target tracking capability and performance.

4.2. Experiment results

Fig. 10 shows the evaluation indexes of the YOLOv8x training model over 50 epochs, including the parameters Precision, Recall, mAP@50, and mAP@50-95. The model's precision increases steadily over the epochs, ranging from about 0.3 to over 0.8 at the end of the training process. This graph shows that the model increasingly predicts more accurately the real objects and people of interest, significantly reducing false positives. Recall also increases steadily, initially at 0.4, reaching about from 0.7 to 0.75 at the end of the epoch. This shows that the ability to fully detect the objects to be recognized is increasingly improved, reducing target omissions. The mAP@50 index (mean average precision at IoU=0.5) remains high, gradually increases over the epochs, and stabilizes around 0.7.

(a) *Experiment results with scenario 1*: In Scenario 1, the target was initially placed at a predefined offset from the image center. The RGB and thermal vision modules first detected the human target and provided the bounding-box center coordinates to the pan-tilt controller. Representative detection and tracking results are shown in Fig. 11. Based on the detected target position, the controller continuously adjusted the pan and tilt angles to minimize the image-plane tracking error.

Fig. 11 shows that when the target moves, the system is able to maintain target detection. The pan error decreases from -674px to 5px , and the system shifts to keep up with the target. From the above results, the authors found that the error indicators displayed on the interface show a significant reduction in control error during tracking, confirming the effectiveness of the pan-tilt control mechanism in re-centering the target.

Fig. 12 compares the simulated and experimental pan-tilt responses for Scenario 1.

Fig. 12(a) shows the controller results with the step motor simulation model in Matlab built based on the mathematical model in equations (11), (12). From the graph, it can be seen that the normalization error of the two pan and tilt axes decreases rapidly over time and gradually stabilizes around the value 0, indicating that the system achieves good stability. The RMSE = 0.0831, MAE = 0.0608 and STD = 0.0568 indices are all low, confirming that the controller maintains small tracking error and stable oscillation during operation.

Fig. 12(b) illustrates the results of the ANFIS controller with the actual model. The results show that the errors of the pan and tilt axes remained around the stable region during the survey period. However, due to the influence of mechanical noise and actual operating conditions on the control quality, the RMSE = 0.1504, MAE = 0.1097, and STD = 0.1029 values increased, leading to higher errors and oscillations compared to the simulation results.

(b) *Experiment results with scenario 2*: The authors conducted a similar experiment to scenario 2 with the goal of continuous movement. In scenario 2, the human target moves continuously along a nonlinear trajectory with varying speed in order to evaluate the dynamic tracking capability of the proposed pan-tilt system. Unlike Scenario 1, where the target is stationary, this experiment requires the controller to continuously update

the pan and tilt commands according to the changing target position obtained from the vision module. The objective is to assess the adaptability, tracking smoothness, and stability of the system under continuous target motion.

Fig. 13 shows the results of stable target detection and tracking on RGB and thermal images. In both cases, the YOLO frame still surrounds the target and the unnormalized control error is reduced.

The corresponding pan-tilt tracking errors are presented in Fig. 14(a) for simulation and Fig. 14(b) for the experimental test. Fig. 14(a) shows that the errors of the two pan and tilt axes decrease gradually with small amplitude and shows that the system maintains good target tracking capability under moving conditions. The RMSE = 0.1031, MAE = 0.0859 and STD = 0.0571 are low, reflecting good tracking quality and stability of the controller in the simulated environment.

Fig. 14(b) shows that under experimental conditions, the pan and tilt axis errors fluctuate significantly more, especially on the pan axis, due to the influence of noise, nonlinear motion, and uncertainties of the real system. The RMSE, MAE, and STD indices are higher than in the simulation but still track the target and give a deviation close to zero.

(c) *Experiment results with scenario 3*: In Scenario 3, the experiment was conducted in a swimming pool to evaluate the system's tracking recovery capability under temporary target loss. The human target moved in the water and was intermittently partially or completely obscured from the camera view due to body submergence, water-surface disturbances, and controlled occlusion. When the target temporarily disappeared, the tracking system maintained the previous target state for a predefined period and continuously searched for its reappearance. Once the target became visible again, the vision module re-detected the person and the pan-tilt controller resumed closed-loop tracking, as shown in Fig. 15. This scenario was used to evaluate target-loss handling, re-detection capability, and tracking recovery before conducting the more challenging real-world coastal experiment in Scenario 4.

Fig. 16 compares the pan-tilt tracking performance in Scenario 3 under simulated and swimming-pool experimental conditions. Fig. 16(a) presents the simulated response, while Fig. 16(b) shows the measured response obtained during the swimming-pool experiment, where temporary target loss and re-detection caused larger tracking fluctuations.

The RMSE = 0.1367, MAE = 0.1085, and STD = 0.0829 are relatively low, confirming that the system maintains fairly good stability during temporary target loss. Fig. 16(b) shows that under experimental conditions, the errors of the pan and tilt axes increase significantly and large oscillations appear when the target is lost from the observation area, especially on the pan axis, reflecting the clear influence of noise and real-world uncertainties.

(d) *Experiment results with scenario 4*: Scenario 4 was conducted in a real coastal environment to evaluate the pan-tilt tracking system under wave-induced disturbances, low illumination, image blur, and intermittent target occlusion. During the experiment, sea

waves partially or completely obscured the person, causing temporary target loss and subsequent reappearance. The system's tracking stability, target re-detection, and recovery capability were evaluated under these realistic operating conditions, as shown in Fig. 17.

Fig. 18 shows that under experimental conditions. The system achieved such as: RMSE = 0.1282, MAE = 0.0950, and STD = 0.0862. Despite several transient error peaks caused by target occlusion and re-detection, the overall tracking error remained relatively low, indicating acceptable stability and recovery performance.

The Scenario 4 experiment was conducted in a real coastal environment; however, the test conditions were limited to a specific beach area, a relatively short observation period, and a restricted range of sea states. Severe weather, larger waves, longer target distances, and highly dynamic illumination conditions were not fully evaluated. Therefore, further sea trials under more diverse environmental and operational conditions are required to comprehensively assess the robustness and generalizability of the proposed system.

The RGB camera is used as the primary sensing device during daytime operation. Under nighttime or very low-illumination conditions, the system switches to the thermal camera to maintain human-target visibility. The thermal image is processed using the same target-detection and coordinate-extraction procedure as the RGB image. After detecting the human target, the bounding-box center is calculated and converted into pan and tilt image-coordinate errors, which are subsequently processed by ByteTrack and the ANFIS controller. Therefore, thermal imaging extends the operating capability of the proposed system to nighttime conditions without changing the downstream tracking and control structure.

4.3. Discussion

This study develops a target tracking model for maritime search and rescue operations. The research results show that the model achieved good stability between simulation and real-world conditions based on RMSE, MAE, and STD indices. The proposed system demonstrated stable target detection and pan-tilt tracking in laboratory, swimming-pool, and real coastal experiments. RGB/thermal cameras and LiDAR provide complementary visual and distance information, improving target localization under varying conditions. However, waves, water reflections, low illumination, partial submergence, sparse LiDAR returns, and actuator delay can still increase tracking errors or cause temporary target loss. Further improvements in sensor fusion, filtering, target prediction, and actuator response are therefore required for reliable operation under more severe sea conditions.

Table 2 summarizes the tracking performance across the four experimental scenarios. As expected, Scenario 3 exhibits the largest errors because of temporary target loss and re-detection, whereas the system maintains lower tracking errors in the static, dynamic, and real-world coastal tests, confirming its ability to operate under different levels of environmental disturbance.

To provide a broader quantitative assessment, Table 3 compares the proposed system with representative

maritime detection and tracking approaches. Since these studies employ different datasets, sensor configurations, target types, and evaluation protocols, the reported metrics are presented according to the original studies and are not interpreted as direct one-to-one performance comparisons.

As shown in Table 3, recent studies generally focus either on maritime object detection or visual tracking, whereas the proposed system combines RGB/thermal detection, LiDAR-assisted sensing, temporary target-loss handling, and closed-loop pan-tilt tracking. Although its detection accuracy is lower than several specialized detection-only models, the proposed system was additionally validated in swimming-pool and real coastal experiments, demonstrating target re-detection and tracking recovery under wave-induced occlusion. This system-level validation represents the main practical contribution of the proposed approach.

5. CONCLUSION

This paper developed a maritime target detection and tracking system integrating YOLOv8, ByteTrack, and an ANFIS-based pan-tilt controller. Experimental results demonstrated stable tracking performance with relatively low errors under static and nonlinear target motion, environmental disturbances, and temporary target loss. The results indicate the practical feasibility of the proposed approach for maritime search and rescue applications.

Despite these promising results, several limitations remain under real-world conditions. The experimental system exhibited actuator delay, mechanical noise, and larger pan-tilt oscillations than those observed in simulation, particularly under environmental disturbances and system uncertainties. Therefore, further improvements are required before deployment under more demanding maritime conditions.

Future work will focus on improving the actuator mechanism, noise filtering, multisensor fusion, and tracking recovery under temporary target loss. Further sea trials onboard the Vietnam-Korea VMU vessel will also be conducted to evaluate long-term stability and tracking accuracy under realistic maritime conditions. In addition, more effective fusion of RGB and thermal imaging will be investigated to improve target detection under nighttime, foggy, and adverse weather conditions.

ACKNOWLEDGMENT

This research is funded by Vietnam Maritime University.

REFERENCES

- [1] M. I. H. Zaidi Farouk, Z. Jamil, and M. F. Abdul Latip, "Towards online surface water quality monitoring technology: A review," *Environ. Res.*, vol. 238, p. 117147, 2023.
- [2] F. Ponzini, D. Van Hamme, and M. Martelli, "Human detection in marine disaster search and rescue scenario: A multi-modal early fusion approach," *Ocean Engineering*, vol. 340, Art. no. 122341, 2025.
- [3] A. Alshibli and Q. A. Memon, "Benchmarking YOLO models for marine search and rescue in variable weather conditions," *Automation*, vol. 6, no. 3, Art. no. 35, 2025.
- [4] M. Pobar, "Detekcija objekata s pomorskih nadzornih kamera," *Politehnika*, vol. 8, no. 1, pp. 42–49, 2024.
- [5] H. Karnan et al., "Visual Servoing of unmanned surface vehicle from small tethered unmanned aerial vehicle," *arXiv*, Cornell Univ., 2022.
- [6] K. Bhaganagar et al., "Design control system for pan-tilt camera for visual tracking based on ADAR method taking into account energy output," *E3S Web Conf.*, vol. 279, pp. 2007, 2021.
- [7] K. Bhaganagar et al., "A novel machine-learning framework with a moving platform for maritime drift calculations," *Front. Mar. Sci.*, vol. 9, 2022.
- [8] F. Lu et al., "Active tracking of moving objects with speed variations using a novel PTZ camera-based machine vision technique," *Elektron. ir Elektrotehnika*, vol. 31, no. 2, pp. 60–70, 2025.
- [9] M. H. Lui et al., "An adaptive YOLO11 framework for the localisation, tracking, and imaging of small aerial targets using a pan-tilt-zoom camera network," *Eng.*, vol. 5, no. 4, pp. 3488–3516, 2024.
- [10] P. C. Tran et al., "Design control system for pan-tilt camera for visual tracking based on ADAR method taking into account energy output," *E3S Web Conf.*, vol. 279, p. 2007, 2021.
- [11] F. Luo et al., "ShipsMOT: A comprehensive benchmark and framework for multiobject tracking of ships," *IET Computer Vision*, 2025.
- [12] A. Hakimi and H. Farjami, "Enhanced significant wave height prediction in the Southern ocean using an ANFIS model optimized with subtractive clustering," *Sci. Rep.*, vol. 15, no. 1, 2025.
- [13] A. Menteş and M. Yetkin, "An application of soft computing techniques to predict dynamic behaviour of mooring systems," *Brodogradnja*, vol. 73, no. 2, pp. 121–137, 2022.
- [14] H. Fakhruddin et al., "Strategi implementasi adaptive neuro fuzzy inference system (ANFIS) pada kendali motor induksi 3 fase metode vektor-tidak langsung," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 9, no. 4, p. 786, 2021.
- [15] O. Hassanein et al., "Unmanned underwater vehicles: applications and challenges," *J. Hunan Univ. Nat. Sci.*, vol. 50, no. 11, 2023.
- [16] M. Rizk et al., "Optimization of deep-learning detection of humans in marine environment on edge devices," *28th IEEE International Conference on Electronics, Circuits, and Systems (ICECS)*, pp. 1–4, 2022.
- [17] Y. Chen et al., "A novel method for water surface debris detection based on YOLOV8 with polarization interference suppression," *Photonics*, vol. 12, no. 6, p. 620, 2025.
- [18] Y. Yuan et al., "Adaptive fuzzy game-based energy-efficient localization in 3D underwater sensor networks," *ACM Trans. Internet Technol.*, vol. 22, no. 2, pp. 1–20, 2021.
- [19] L. Duan et al., "Research on seismic signal denoising model based on DnCNN network," *Applied Sciences*, vol. 15, no. 4, 2025.
- [20] Y. Wang and V. Y. Mariano, "A multi object tracking framework based on YOLOv8s and bytetrack algorithm," *IEEE Access*, vol. 12, pp. 120711–120719, 2024.
- [21] S. Yun et al., "Action-decision networks for visual tracking with deep reinforcement learning," in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1349–1358, 2017.
- [22] M. Doloiu et al., "Modeling and control of a pan-tilt servo system for face tracking using deep learning and PID," *Eng. Proc.*, vol. 113, p. 75, 2025.

- [23] Y. Sun, "Analysis for center deviation of circular target under perspective projection," *Eng. Comput.*, vol. 36, no. 7, pp. 2403–2413, 2019.
- [24] W. Choi and S. Savarese, "Multiple target tracking in world coordinate with single, minimally calibrated camera BT – Computer Vision – ECCV 2010," K. Daniilidis, P. Maragos, and N. Paragios, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 553–567, 2010.
- [25] H. M. Lee and W. C. Choi, "Algorithm of 3D spatial coordinates measurement using a camera image," *J. Image Graph.*, vol. 3, no. 1, pp. 30–33, 2015.
- [26] D. Xu et al., "EMA-YOLO: A Novel target-detection algorithm for immature yellow peach based on YOLOv8," *Sensors*, vol. 24, p. 3783, 2024.
- [27] P. Singh, R. Gottumukkala, and A. Maida, "An analysis of kalman filter based object tracking methods for fast-moving tiny objects," 2025.
- [28] D. Mlakić, H. R. Baghaee, and S. Nikolovski, "A novel ANFIS-based islanding detection for inverter-interfaced microgrids," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4411–4424, 2019.
- [29] P. Mehta and M. Jain, "ANFIS and metaheuristics for double orbit retrieval queue with complete and working vacations," *Appl. Soft Comput.*, vol. 167, p. 112473, 2024.
- [30] D. Joshi, A. Mishra, and S. Anand, "ANFIS based knee angle prediction: An approach to design speed adaptive contra lateral controlled AK prosthesis," *Appl. Soft Comput.*, vol. 11, no. 8, pp. 4757–4765, 2011.
- [31] C. Volosencu, *Adaptive Neuro-Fuzzy Inference System as a Universal Estimator*, 2024.
- [32] M. A. AbouElaz et al., "ANFIS-optimized control for resilient and efficient supply chain performance in smart manufacturing," *Results Eng.*, vol. 25, pp. 104262, 2025.
- [33] H.-C. Liao et al., "Automatic zooming mechanism for capturing clear moving object image using high definition fixed camera," in *2017 19th International Conference on Advanced Communication Technology (ICACT)*, pp. 869–876, 2017.
- [34] N. AL-Rousan, N. A. Mat Isa, and M. K. Mat Desa, "Efficient single and dual axis solar tracking system controllers based on adaptive neural fuzzy inference system," *J. King Saud Univ. – Eng. Sci.*, vol. 32, no. 7, pp. 459–469, 2020.
- [35] M. N. S. Jagirdar et al., "Implementation of adaptive neuro-fuzzy inference system in design and process optimization of latent heat storage system," *J. Energy Storage*, vol. 89, p. 111810, 2024.
- [36] X. Chen et al., "Augmented ship tracking under occlusion conditions from maritime surveillance videos," *IEEE Access*, vol. 8, pp. 42884–42897, 2020.
- [37] F. Ponzini, D. Van Hamme, and M. Martelli, "Human detection in marine disaster search and rescue scenario: A multi-modal early fusion approach," *Ocean Engineering*, vol. 340, Art. no. 122341, 2025.
- [38] Z. Jin et al., "MES-YOLO: An efficient lightweight maritime search and rescue object detection algorithm with improved feature fusion pyramid network," *Journal of Visual Communication and Image Representation*, 2025, Art. no. 104453.
- [39] Z. Gao et al., "HPRT-YOLO: Enhanced lightweight UAV maritime search and rescue model based on YOLOv11n," *iScience*, vol. 29, no. 4, Art. no. 115366, 2026.
- [40] B. Zhao et al., "Enhanced YOLO11 for lightweight and accurate drone-based maritime search and rescue object detection," *PLOS ONE*, vol. 20, no. 7, e0321920, 2025.