Intelligent Human Systems Integration (IHSI), Vol. 200, 2026, 395–401
https://doi.org/10.54941/ahfe1007098

Improving the Accuracy of Automatic Object Tracking by Using Posture Recognition

Satoru Inoue[1], Mark Brown[2], and Tomofumi Yamada[3]

[1]Surveillance and Communication Department, Electronic Navigation Research Institute, Chofu, Tokyo 182-0012, Japan
[2]Air Traffic Management Department, Electronic Navigation Research Institute, Chofu, Tokyo 182-0012, Japan
[3]Fixstars Corporation, Minato-ku, Tokyo, 108-0023, Japan

ABSTRACT

This paper concerns an automatic object tracking system for Point-Tilt-Zoom (PTZ) cameras using image recognition technology, aimed at a new type of airport air traffic control system known as Remote Tower. Airport air traffic controllers work by visual observation of the airport and its vicinity from a glass-walled room at the top of a control tower. In Remote Tower, a "panoramic" image from cameras at the airport replaces the out-of-the-window view, allowing the physical building to be eliminated and the control service to be moved away the airport. The PTZ camera serves the role of binoculars, allowing close-up views of aircraft both moving on the airport surface and flying in the vicinity. Automatic PTZ tracking of selected objects relieves operator workload, However, tracking based solely on image recognition processing of PTZ video frames presents challenges, such as tracking switching to an unintended target when multiple aircraft appear within the frame, and continuity of tracking when the view of target aircraft is temporarily occluded by obstacles. To address such situations, we investigated using object orientation in addition to object identification. Preliminary trials confirmed improved tracking stability and resistance to target switching for slow-moving aircraft that cross in the PTZ view. This study systematises the control challenges of Automatic Object Tracking using image recognition and discusses potential solutions to these issues.

Keywords: Automatic object tracking, Image recognition, Remote tower technologies

INTRODUCTION

In recent years, in addition to classical machine vision algorithms, Artificial Intelligence (AI)-based techniques such as image segmentation and object detection and recognition have been utilised in the transport sector for various purposes (Zhang, 2025). Object detection for collision avoidance is used in both driver assistance or autonomous "self-driving" using on-board cameras(Turay, 2022), and in traffic monitoring and management using roadside cameras (Dilek, 2023). One application in the field of aviation is "Remote Tower" (Fürstenau, 2022). Airport air traffic controllers work by observing traffic on the airport surface and flying in the vicinity from a glass-walled

Received October 10, 2025; Revised November 28, 2025; Accepted December 13, 2025; Available online February 1, 2026

room on top of a control tower building. Remote Tower uses video cameras to replace the out-of-the-window view with a wide field-of-view "panorama" image from video cameras on top of a mast and allowing controllers to provide services from a remote location. Digitization of the image streams also presents an opportunity to develop functions to support controller tasks and improve situational awareness.

Airport air traffic controllers must maintain a continuous visual watch, not only of aircraft and vehicles on the airport surface, which are slow-moving and easily visible, but also of aircraft flying within a few kilometers of the airport, which are harder to find and keep track of. Controllers sometimes desire a close-up view of a specific aircraft or vehicle of interest. In a conventional control tower, they would use binoculars, and in a Remote Tower, this facility is often provided using a pan-tilt-zoom (PTZ) camera. Digitilazation allows the introduction of an automatic PTZ tracking feature to enable the controller to share attention with other tasks while monitoring the PTZ image, something not possible with binoculars.

We have developed a PTZ tracking function, an overview of which is shown in Figure 1. The tracking function uses two inputs: "optical", based on recognising and tracking aircraft or ground vehicles in the PTZ camera image, and "sensor", which uses position and height information from a surveillance sensor such as a radar, multilateration (a triangulation system using radio signals from transponders equipped in each vehicle), or position information broadcast by the vehicles themselves (ADS-B). In use, the air traffic controller selects the target aircraft or vehicle from a list of aircraft that are being tracked by the airport surveillance system. The position and height information from the surveillance system is then used to cue to PTZ camera to point in the target's direction. Subsequently, the system utilises image recognition to locate objects within the camera's field of view that match the desired target's class (aircraft, vehicle), and then continuously and smoothly adjusts the PTZ direction to keep the target close the centre of the image. This paper discusses the challenges and solutions relating to this function.

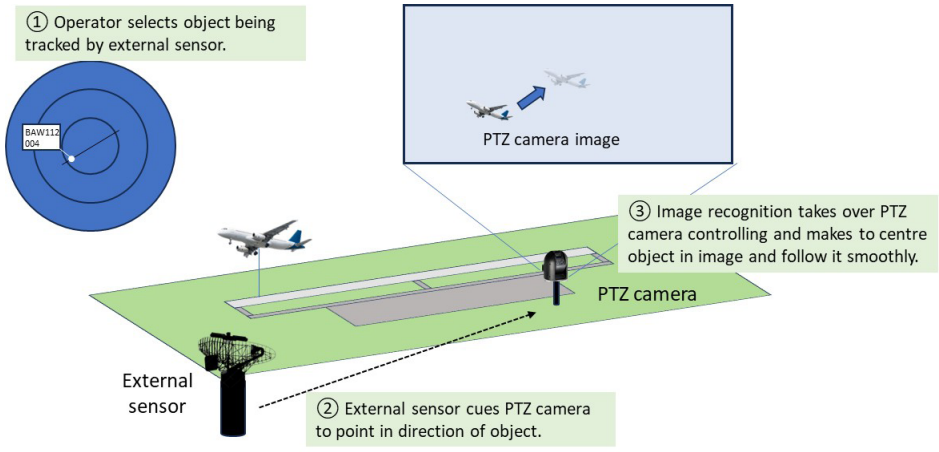

Figure 1: Overview of the automatic tracking system using PTZ cameras.

CHALLENGES IN AUTOMATIC OBJECT TRACKING USING IMAGE RECOGNITION

We developed the PTZ tracking function based on an earlier-developed function for recognising aircraft and vehicles at airports and their surroundings in video images using the You Only Look Once (YOLO) convolutional neural network-based object detection framework; specifically, YOLOv8. Traditional methods for recognising arbitrary targets in images require a two-step process: extraction of portions of the image that may contain target objects (segmentation), then searching the candidate regions to identify target objects and determine their precise positions. YOLO, on the other hand, has the advantage of carrying out segmentation and target extraction and recognition in a single step (as its name 'you only look once' implies), enabling rapid processing. Segments are extracted as rectangular bounding boxes that contain objects of recognized classes, along with a confidence value.

In our initial implementation of the tracking function, the centre of the detected segment's bounding box was calculated as an approximate target position. The PTZ pointing controller then used changes in this position over several frames to estimate the direction and speed of its apparent motion (velocity vector). Although this approach is simple to implement, it has issues caused by the fundamental problem of not being able to recognize specific objects: YOLO simply detects object classes in each frame and cannot detect a specific previously-observed instance of a given object class. This leads to target tracking "switching". While tracking one object, if another object of the same object class but with a higher confidence value enters the frame, tracking may shift to following that object instead. A similar issue occurs when an object being tracked is temporarily lost from view (e.g., if hidden by an obstacle). The object's velocity vector may be used to continue to move the PTZ open-loop based on the last known velocity vector and hoping to re-acquire the target, but if another vehicle of the same object class enters the view, it may be mistaken for the target object.

As our first attempt to suppress this switching phenomenon, we attempted to discriminate the target from other objects by comparing their current movements with the past movement of the target object. A polynomial function is estimated that interpolates between two linear vectors derived from the "last position of the past movement" and the "velocity of the past movement", and the "first position of the current movement" and the "velocity of the current movement", which are calculated from segment bounding boxes as described above. Current and future movements are judged to be closer together when the coefficients for the cubic "change in acceleration" and the quadratic "acceleration" terms of the interpolating polynomial are smaller.

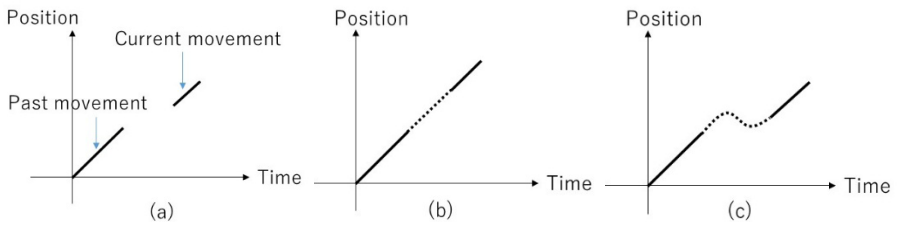

Figure 2: Basic idea of estimation of movement from velocity data.

Figure 2 illustrates the concept of the method. Fig. 2(a) shows the past movement of the target object being tracked and the current movement of a candidate target. If they can be connected almost linearly as shown in Fig. 2(b), the values of the quadratic and cubic terms of the interpolating polynomial will be small. On the other hand, if the segments are significantly linearly disjointed as shown in Fig. 2(c), the cubic coefficient will be large. The candidate with the smallest interpolating polynomial coefficient values is therefore judged to be the target being tracked. We evaluated the method using out testbed system with live and replay video from our Sendai airport branch cameras. Although the principle was found to be valid, when aircraft crossed at slow speeds, large frame-to-frame fluctuations in segment bounding boxes tended to occur, preventing reliable estimation of velocity vectors. Furthermore, when object detection was intermittent from frame to frame due to temporary occlusion, resulting in gaps in the data, the interpolating polynomials became smoother, leading to instances where other targets were misidentified. Figure 3 shows a PTZ image at the time of an impending switch between a tracked target, the aircraft at the centre of the frame facing away from the camera which is moving diagonally towards the upper right, and the left-facing aircraft visible to the upper left of the target which has moved from the right.

Figure 3: Example situation where a switch occurred.

Figures 4 shows the time histories of the detected movement across the view of the true target (blue trace) and switched target an example of the occurrence of a 'Switch'. The vertical axis shows the azimuth angle of the PTZ camera, and the horizontal axis shows time t. (The azimuth zero value and time $t = 0$ values are not important here.) The aircraft "cross" paths between $t = 10$s and $t = 22$s, indicated by vertical blue lines. At around $t = 22$ seconds, the detection of the target aircraft became somewhat unstable for about 1s, shown by red circle. Consequently, the movement of the "switched target" aircraft became more closely correlated with the past movements of the target than the actual target, leading to a switch in tracking (black arrow).

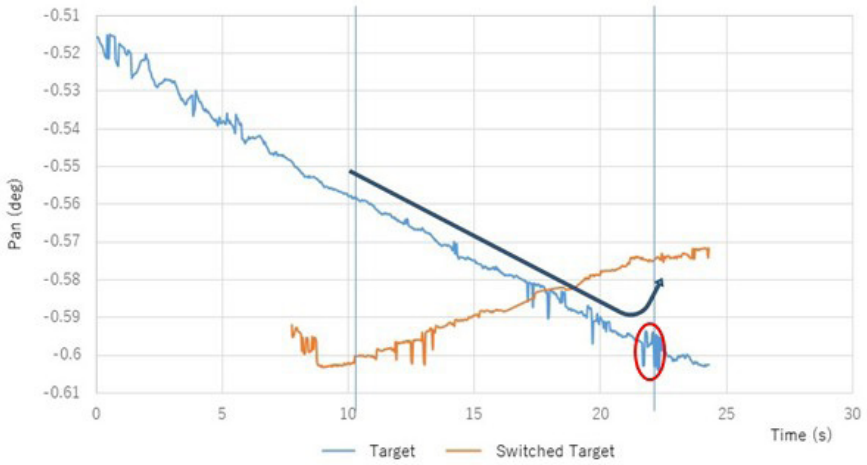

Figure 4: State of the coordinates where the switch occurred in Figure 3.

Measures to address this issue necessitates careful consideration of trade-offs. For example, averaging unstable states might resolve the problem in the scenario above, but may reduce the reliability of tracking during manoeuvres such as sharp turns. When an object is moving slowly, attempting to derive velocity vector from its image segmentation bounding boxes proves problematic as frame-to-frame fluctuations in the bounding box introduces noise. This highlights the necessity for techniques specifically designed to address such issues.

OBJECT IDENTIFICATION BASED ON ORIENTATION

We investigated whether continuous tracking of target aircraft could be achieved even in crossing situations by using aircraft orientation obtained using the "Pose" feature of YOLOv8. As stated above, when an object moves across the video image at low speed, fluctuations in the segmentation bounding box lead to fluctuations in the object's estimated position, introducing noise into its velocity vector. This creates a situation where track using object detection alone becomes difficult. Therefore, we investigated utilising object orientation information as an additional detection parameter by using the YOLO 'Pose' feature to recognise aircraft orientation.

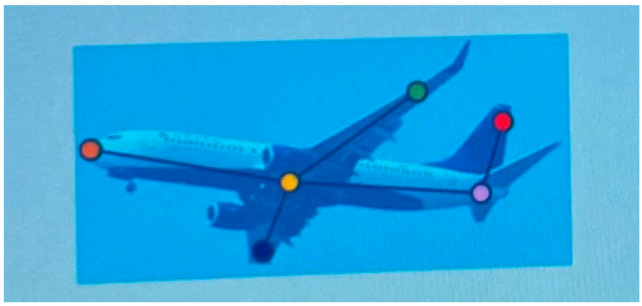

Figure 5: Example of annotation data creation for aircraft attitude recognition.

Orientation recognition was trained by annotating a set of aircraft images with six points—nose, fuselage, right wing, left wing, tail, and vertical tail fin—as shown in Figure 5. The connection relationships between each position—such as the nose and body, or the tail and tail wing tip—were also defined to enable recognition of component parts even when only a portion of the aircraft is visible. Approximately 20,000 annotation data points for orientation recognition were created from daytime images.

Figure 6 illustrates an example of orientation recognition from the image of an aircraft against a complex airport background. Although the aircraft's right wing is not visible, its orientation is recognised from the remaining visible components comprising the nose and body, the body and left wing tip, the body and tail, and the tail and vertical tail wing tip. Using "pose" information, a mechanism has been constructed to estimate the direction in which an aircraft is facing.

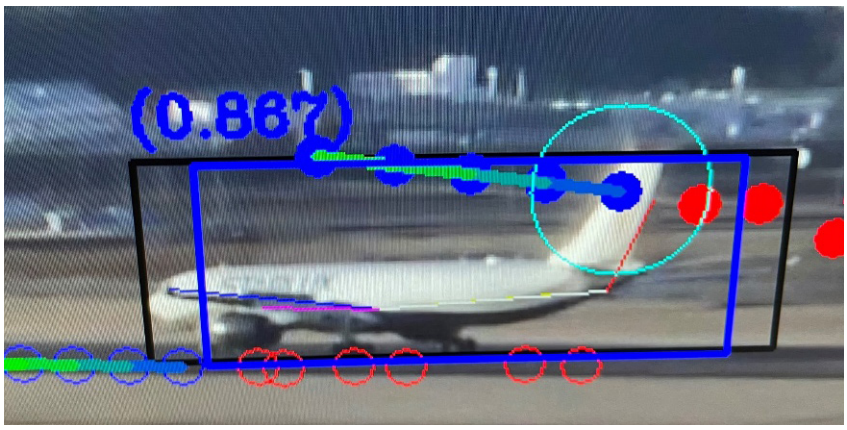

Figure 6: Example of aircraft posture recognition and target plane tracking.

We applied the orientation recognition to improve the accuracy of the velocity vector estimate. Instead of the centre of the object's segment bounding box, which can be unstable at low speeds, we utilised the tip of the tail fin obtained from posture recognition as the tracking point. The reasons for selecting the tail fin tip are that it is visible regardless of the direction the aircraft is facing, and it is less likely to be occluded by other aircraft and structures on the airport surface, giving more stable tracking. In preliminary test using video images from Sendai airport, using the tail fin tip reduced fluctuations in object position compared to the previous method, particularly in the PTZ azimuth direction (i.e. horizontally across the image frame). By utilising two parameters derived from image recognition - the direction derived from the aircraft's orientation and the velocity vector derived from the tracking point - a mechanism has been established that is resilient to situations where tracking would previously have "switched" from the intended target to another object.

Moreover, by recognising aircraft attitude based on the component parts determined from visible reference points - such as the nose and body, or the left wing tip and body - it is also possible to determine when the view of

the aircraft becomes obstructed. For example, if the initial visible nose-body section cannot be identified, and the number of component parts that cannot be identified increases with time, it can be determined that the aircraft has become hidden. This will enable better judgement of the timing to switch to PTZ control from closed-loop tracking to open-loop "coasting" based on estimated velocity vector, or from optical to surveillance sensor mode.

On the other hand, it is anticipated that using orientation information for tracking may prove difficult in the case when multiple aircraft facing in the same direction such as traffic jam situations on the taxi way are concentrated in the image frame.

CONCLUSION

In the case study, we could determine the aircraft's orientation through image recognition. By utilising this orientation recognition result as a determination parameter for tracking, we confirmed that continuous tracking is achievable even in situations where automatic target tracking would previously be impossible using only velocity vectors.

REFERENCES

Dilek, Esma, Dener, Murat. (2023) Computer vision applications in intelligent transportation systems: a survey, Sensors 23 (6), 2938, https://doi.org/10.3390/s23062938

Fürstenau, Norbert, ed. (2022) Virtual and Remote Control Tower. Research, Design, Development, Validation, and Implementation, Springer Nature.

Jocher, G., Chaurasia, A. and Qiu, J. (2023) YOLO by Ultralytics, https://docs.ultralytics.com/models/yolov8/

Turay, Tolga, Vladimirova, Tanya. (2022) Towards Performing Image Classification and Object Detection with Convolutional Neural Networks in Autonomous Driving Systems: A Survey, IEEE Access 10(1):1-1, DOI:10.1109/ACCESS.2022.3147495

Zhang, Jingyu. Cao, Jin. Chang, Jinghao and Li, Xinjin, (2025). Research on the Application of Computer Vision Based on Deep Learning in Autonomous Driving Technology, Proceedings of the 2023 International Conference on Wireless Communications, Networking and Applications (pp. 82–91), DOI:10.1007/978-981-96-2409-6_9