Dual-Camera Active Acquisition for Automated Small-Object Dataset Construction

Jackie Wang*, JC Vaught†, Douglas Cahl‡, Yi Wang§
Department of Mechanical Engineering, University of South Carolina, Columbia, SC, 29201

Deep learning performance on small objects is frequently bottlenecked by the quality and quantity of training data rather than model architecture. To address this, we propose an active dual-camera system for automated small-object dataset generation, specifically designed to overcome the resolution limits of static wide-angle surveillance. The system leverages a fixed wide-angle camera for target discovery and a PTZ unit for detailed interrogation. The control framework transitions from open-loop predictive slewing to closed-loop visual tracking, compensating for mechanical latencies and slow zoom mechanics. Implemented on an NVIDIA Jetson Orin, the system runs concurrent detector instances, one on the GPU and one on the Deep Learning Accelerator (DLA) to achieve 30 fps throughput. High-resolution object verifications are projected back into the wide frame using a static homography-based calibration, creating high-confidence labels for targets that appear as only a few pixels in the wide view. We validate the system design against a test case of distant aircraft in daylight, analyzing the trade-offs between slew speed (120°**/s) and zoom settling time. Preliminary analysis suggests this active acquisition paradigm can improve label precision significantly and reduce the human effort required for small-object dataset curation.**

I. Introduction

While recent surveys [1, 4] have demonstrated significant architectural improvements in generic object detection, they have failed to fully solve a problem that has persisted for decades: small object detection. The core issue remains one of pixel-level information loss, since no amount of digital zoom can recover details that were never sampled. Although some super-resolution algorithms show promise by recovering detail via temporal information, these methods are often computationally intensive and ill-suited for strict real-time inference [2].

Even with high-resolution imagery, real-time detection pipelines typically downsample inputs to manageable resolutions (e.g., 640×640) to satisfy compute constraints. This reduction inherently compresses distant targets into featureless blobs (Fig. 1) that are difficult to classify without relying on temporal context (as humans do) or additional sensor modalities [3].

Figure 1 illustrates the challenge in distant aerial surveillance, where a target may occupy fewer than 15×15 pixels in a wide-angle feed. At this resolution, distractors (i.e. birds, cloud edges, specular highlights, or sensor noise) are indistinguishable from the target of interest. Standard digital zooming (cropping) only magnifies these ambiguities [5]. To achieve robust detection, a dataset must contain examples where these confusing cases are resolved, yet human labelers often cannot distinguish them in raw wide-angle footage.

We address this by replacing the human labeler with a dual-camera system designed to automate the construction of small-object datasets. A fixed wide-angle camera provides persistent coverage, while a controllable Pan-Tilt-Zoom (PTZ) camera provides resolution on demand. The key technical challenge lies in the handover, since the system must move a mechanical lens to capture a moving target based on delayed visual data. This system compensates for the ≈ 150 ms latency between sensor capture and motor response and achieves high movement rates (120°/s) without inducing motion blur.

The contributions of this work are presented as follows: i) a hardware-software architecture that synchronizes wide-angle search with narrow-angle verification; ii) a predictive control formulation that compensates for system latency to enable reliable active target acquisition from wide-angle cues; and iii) a validated pipeline for label transfer,

*Graduate Student, Department of Mechanical Engineering, AIAA Student Member.
†Graduate Student, Department of Mechanical Engineering, AIAA Student Member.
‡Professor, Department of Mechanical Engineering.
§Professor, Department of Mechanical Engineering.

Fig. 1 Comparison of Wide-Angle Crop (Digital Zoom), Super-Resolution, and true Optical Zoom active acquisition.

showing how distracting objects rejected by the PTZ can be automatically added to the wide-angle dataset to reduce false positives in future deployment.

II. Related Work

A. Small Object Detection and Resolution Limits

The fundamental bottleneck in small object detection is the scarcity of distinguishing features. When a target occupies few pixels (i.e. 15×15 pixels), class-defining details are often lost entirely [1]. Classical solutions involving super-resolution (SR) attempt to reconstruct this lost detail [2].

However, reliance on SR for scientific data collection is flawed. First, SR is fundamentally generative; it estimates high-frequency details based on learned priors, creating a risk of hallucination where the model reinforces its own biases [5]. Second, the latency advantage of SR is negligible for high-quality restoration. While lightweight models can run in $<$ 30ms on edge accelerators (e.g., Jetson Orin), high-fidelity generative models required for scientific validity often require $>$ 300ms per frame [18]. This exceeds the mechanical slew-and-settle time of our system ($\approx$ 150ms), which is competitive with high-end commercial PTZ units (typically 60–200 ms command latency [16]). Our system therefore chooses the mechanical penalty to obtain optical ground truth rather than the computational penalty for hallucinated details.

B. Active Acquisition vs. Continuous Tracking

Most PTZ tracking literature focuses on the control problem of keeping a target centered in the frame [7, 8]. This requires mitigating total system latency (video encoding + network + mechanical response), which for IP-based systems frequently ranges from 200–500 ms [17]. Our work addresses active acquisition, or "slew-to-classification." Unlike continuous tracking, where the objective is persistence, our objective is information gain via discrete or one-off spot-checks.

Existing active perception systems like VIGIA-E [11] typically optimize for broad area coverage or anomaly detection. Our system instead functions as a sparse query mechanism. It identifies specific low-confidence candidates in the wide field and commits the PTZ resource to verifying them individually. This shifts the challenge from long-term stabilization to fast, accurate separate-and-verify maneuvers.

C. Sensor-Driven Labeling

Reducing manual annotation is a central goal of both semi-supervised learning and active learning. Pseudo-labeling methods such as ASTOD [12] attempt to retrain models using high-confidence predictions, but this approach often fails in the small-object regime where the detector is consistently uncertain [13]. Similarly, active learning strategies like PPAL [15] identify informative samples but still require a human loop [14].

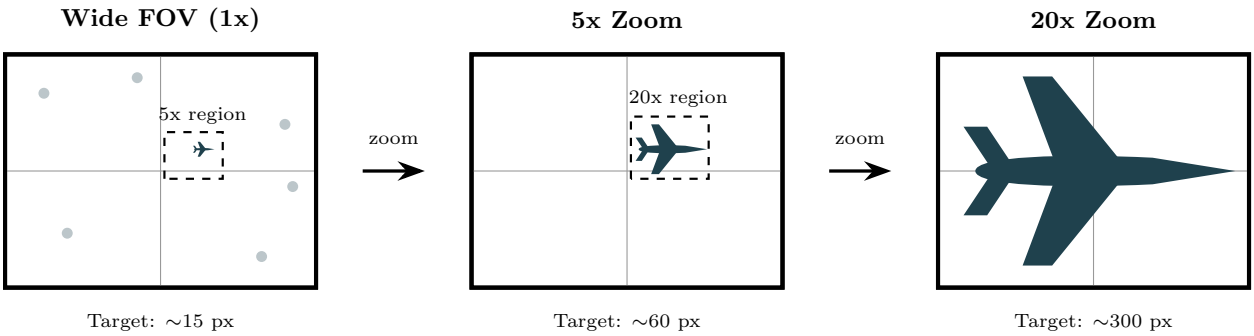

Fig. 2 Visual progression of an active acquisition sequence. A 15-pixel target in the wide field (1x) is indistinguishable from noise. The system cues the PTZ to an intermediate 5x zoom for acquisition, then a 20x zoom for final detailed verification, revealing the aircraft structure clearly.

Our proposed "Active Acquisition" creates a fully automated hybrid. We use the selection logic of active learning (targeting less confident samples) but satisfy the label query using the PTZ camera instead of a human. The success of this automated verification relies on the additional information provided by optical zoom. While the target is ambiguous at 15×15 pixels, the zoomed view restores it to a regime (e.g., $> 100 \times 100$ pixels) where off-the-shelf detectors already achieve near-perfect accuracy [6]. By physically bridging the gap between the surveillance view and the high-resolution training distribution of standard models, we convert a difficult "small object" inference problem into a trivial classification task, enabling the generation of verified small object ground truth labels at scale.

III. System Overview

A. Hardware and Software Roles

The system uses two cameras mounted with a fixed relative pose (Fig. 3):

(1) a fixed wide-angle camera providing continuous coverage and running a real-time YOLO-family detector;

(2) a PTZ camera whose pan, tilt, and zoom can be commanded via a control interface and whose telemetry is logged with timestamps.

The wide camera produces candidate detections and tracklets. A scheduling policy selects targets and commands the PTZ to re-center and zoom. The PTZ stream is analyzed to confirm class and refine bounding boxes. High-confidence PTZ detections are transferred back to the wide-angle frame to create dataset labels. A curation module filters low-quality frames and controls redundancy.

B. End-to-End Pipeline Diagram

IV. Problem Formulation

Let I_t^w be the wide-angle frame at time t, and let I_t^p be the PTZ frame captured at time t with telemetry $\tau_t = (\theta_t, \phi_t, z_t)$ representing pan, tilt, and zoom. The wide detector produces candidate detections $D_t = \{(b_{t,i}^w, s_{t,i}^w, c_{t,i})\}_i$ with bounding boxes $b_{t,i}^w$, confidence scores $s_{t,i}^w$, and class labels $c_{t,i}$.

The system chooses actions a_t that either keep scanning (no PTZ) or command the PTZ to acquire a zoomed observation centered on a selected candidate. The goal is to maximize the expected dataset value subject to control and compute constraints:

$$\max_{\pi} \ \mathbb{E}\left[\sum_{t} U\left(I_t^w, I_t^p, \tau_t\right)\right] \tag{1}$$

subject to PTZ dynamics, latency, bandwidth, and storage budgets. Here $U(\cdot)$ is an acquisition utility that increases with (i) label correctness probability, (ii) localization quality, (iii) novelty/diversity, and (iv) usefulness for training (hard examples and hard negatives).

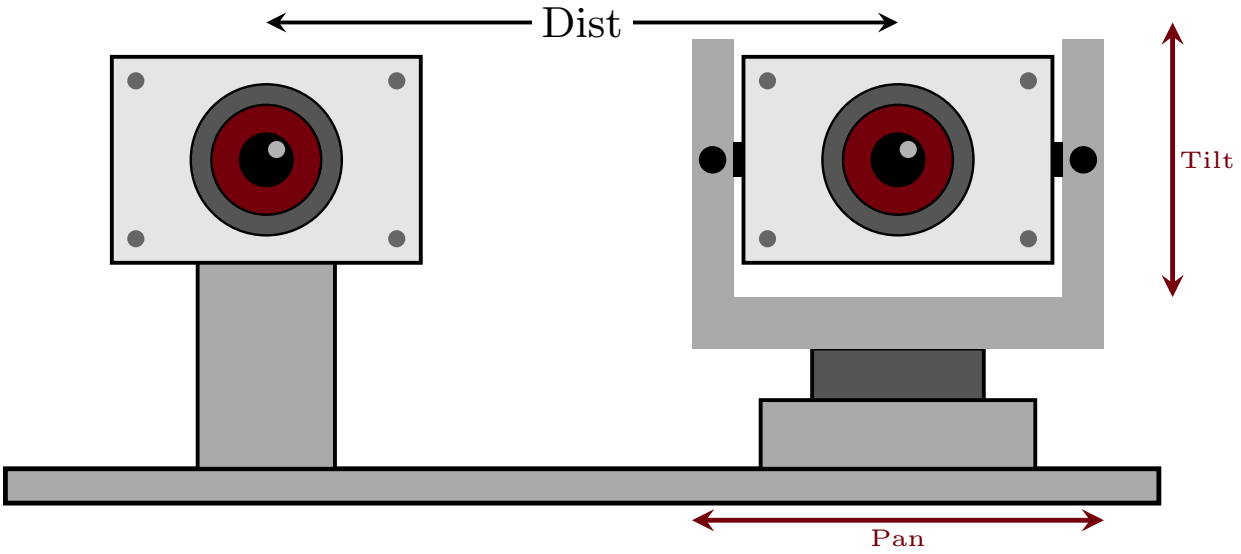

Fig. 3 Physical hardware configuration. A fixed wide-angle camera provides persistent coverage while a 2-axis PTZ camera can be commanded to lock onto and zoom in on candidate targets.

V. Calibration and Cross-View Geometry

A. Camera Model

Both cameras are modeled with intrinsics (K_w, K_p) and distortion parameters. Let $\Pi(\cdot)$ be perspective projection with distortion, and let R_{pw}, t_{pw} map 3D points from wide-camera coordinates to PTZ-camera coordinates.

For sky targets at long range, parallax between cameras is often negligible if the baseline is small relative to range. In that regime, direction-only transfer is effective: rays from each camera correspond to the same world direction, and cross-view mapping can be performed using ray directions on the unit sphere rather than estimated depth.

B. Mapping Wide Detections to PTZ Pan/Tilt Commands

Let (u, v) be the center of a wide detection box b^w in pixel coordinates. After undistortion, the bearing direction in the wide camera frame is

$$\hat{d}_w = \frac{K_w^{-1}[u\ v\ 1]^\top}{\|K_w^{-1}[u\ v\ 1]^\top\|}. \tag{2}$$

Transforming to the PTZ base frame gives $\hat{d}_p = R_{pw}\hat{d}_w$. Using a conventional yaw–pitch parameterization, the commanded pan (yaw) and tilt (pitch) are

$$\theta = \text{atan2}(\hat{d}_{p,y}, \hat{d}_{p,x}), \quad \phi = \text{atan2}(\hat{d}_{p,z}, \sqrt{\hat{d}_{p,x}^2 + \hat{d}_{p,y}^2}). \tag{3}$$

Because PTZ motion and video pipelines have latency, $\hat{d}_p$ is preferably computed from a short-horizon prediction of target motion rather than the instantaneous detection center.

C. Zoom Selection by Target Angular Size

Let h^w be the detected box height in wide pixels. Under small-angle approximation, the apparent angular height is approximately $\alpha \approx h^w / f_w$, where f_w is the wide focal length in pixels. To obtain a desired PTZ pixel height h^p_{des}, select a PTZ focal length $f_p(z)$ such that

$$h^p_{\text{des}} \approx \alpha f_p(z) \approx \frac{h^w}{f_w} f_p(z), \tag{4}$$

then choose the zoom z whose calibrated $f_p(z)$ best matches the required value, clipped to PTZ limits. This approach avoids explicit range estimation and is well suited for distant aircraft.

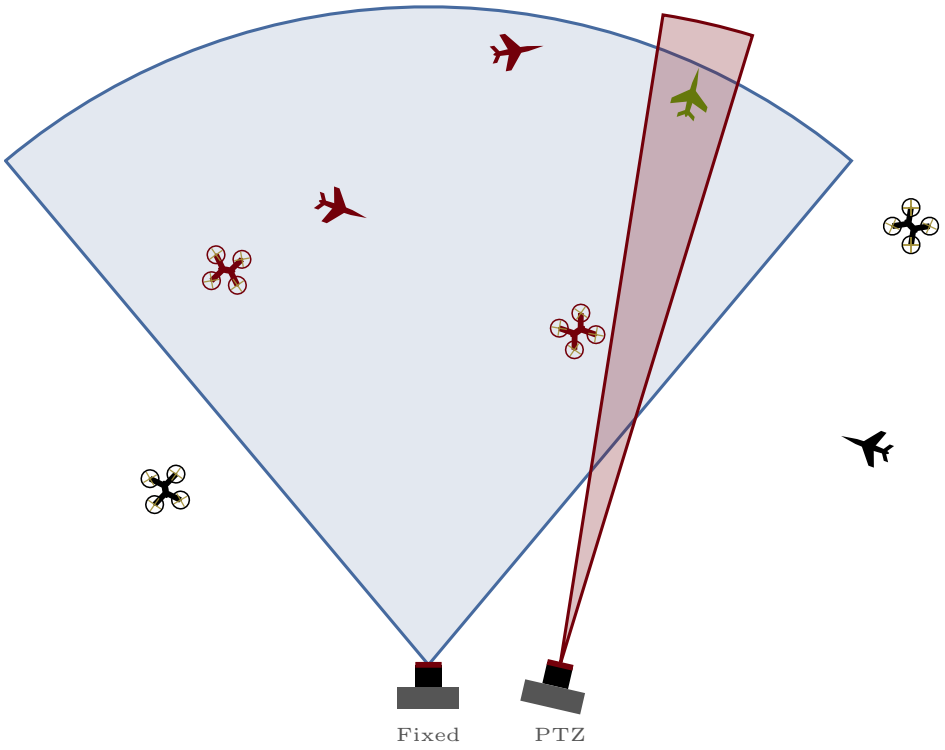

Fig. 4 Field of view comparison. The wide camera (blue) covers approximately 90° for persistent surveillance, while the PTZ camera (red) provides a narrow, steerable, high-resolution view that can be directed to any point within the wide FOV.

VI. PTZ Triggering, Tracking, and Scheduling

PTZ tracking is a coupled perception-and-control problem with dynamic imaging conditions and control delays. Prior PTZ tracking evaluations emphasize that camera motion, latency, and re-centering quality dominate performance differences in practice.

We use wide-camera tracklets to provide temporal coherence and to predict target bearing during PTZ motion. A lightweight predictor (e.g., constant angular velocity with α–β filtering or a Kalman filter) provides a predicted bearing $\hat{d}_w(t + \Delta)$ given command latency Δ, consistent with classical pan/tilt tracking formulations.

A. Utility Function for Triggering and Redundancy Control

The trigger utility balances value and cost. A practical form is a weighted sum of terms:

$$U(o) = \lambda_1 \,\mathrm{Unc} + \lambda_2 \,\mathrm{SizeGain} + \lambda_3 \,\mathrm{Novelty} - \lambda_4 \,\mathrm{Cost}, \tag{5}$$

where Unc increases when the wide detector is unsure (so PTZ confirmation is valuable), SizeGain estimates how much the PTZ will increase target pixels, Novelty reduces redundancy by down-weighting near-duplicates, and Cost captures PTZ time-on-target, slew distance, and opportunity cost (only one PTZ target at a time).

This framing is consistent with object-detection active learning literature that combines uncertainty and diversity, although here the action is a sensor query rather than a human label request.

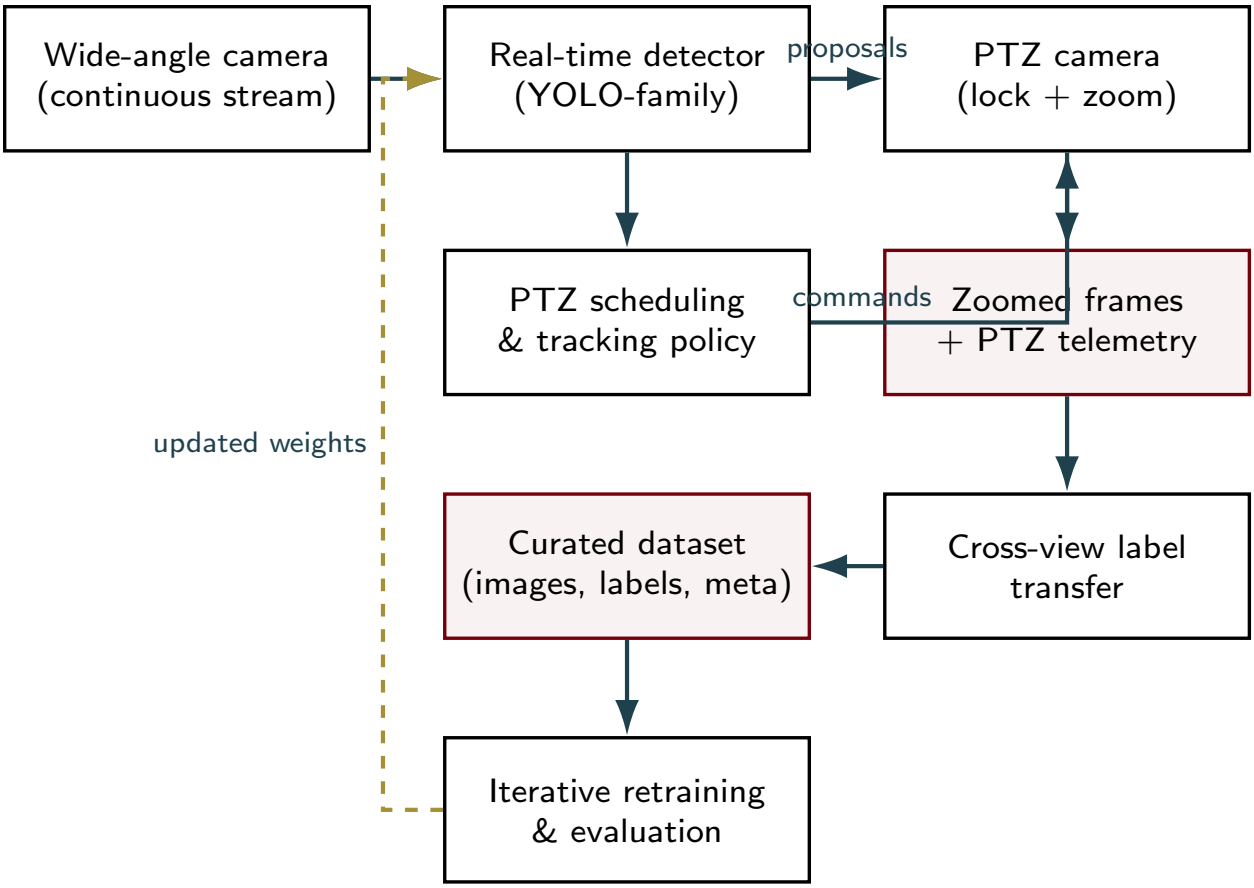

Fig. 5 Dual-camera active acquisition loop. The PTZ camera is triggered by wide-angle detections to improve evidence quality; high-confidence PTZ detections are transferred back to wide-angle frames.

VII. Automated Dataset Construction

A. Label Sources and Cross-View Transfer

The PTZ view is treated as a higher-fidelity label source when it produces a confirmed detection for the target class. The label transfer module maps the PTZ detection back into the wide image coordinate system.

For distant targets, direction-only transfer proceeds by converting the PTZ bounding box corners into bearing rays, rotating those rays into the wide camera frame, and projecting them into the wide image plane. Let (u_k^p, v_k^p) be PTZ box corners; compute PTZ rays $\hat{d}_k^p$ from K_p^{-1}, rotate to wide frame $\hat{d}_k^w = R_{wp}\hat{d}_k^p$, then project:

$$[u_k^w, v_k^w, 1]^\top \propto K_w \hat{d}_k^w, \quad k \in \{1, 2, 3, 4\}. \tag{6}$$

The transferred wide box is the tight axis-aligned rectangle enclosing the projected corners. This procedure uses only calibrated intrinsics and the relative rotation, plus the PTZ telemetry that defines the effective PTZ optical axis at capture time.

B. Quality Gates and Drift Prevention

Self-training and pseudo-labeling can suffer from confirmation bias if low-quality pseudo-labels are admitted. This is widely recognized in semi-supervised detection; adaptive thresholding and robust pseudo-label selection reduce manual tuning and improve stability.

In this system, the primary drift control mechanism is the physical zoom verification: many ambiguous wide detections become unambiguous at higher resolution. Remaining failure modes are handled by quality gates (Fig. 7) that reject samples when motion blur is excessive, exposure is saturated (e.g., sun glare), the target is at the frame boundary,

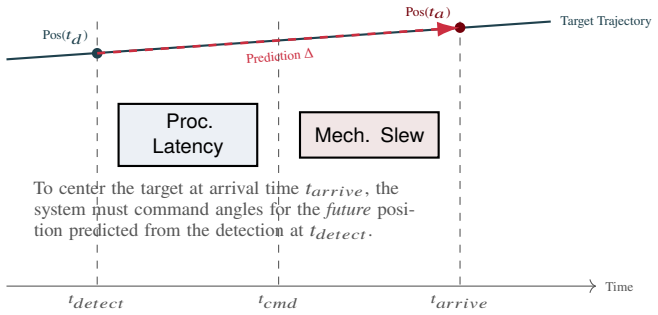

Fig. 6 Latency Compensation. The PTZ command must account for processing delays and mechanical slew time. The target position is extrapolated Δ seconds into the future to ensure the PTZ arrives at the correct bearing.

Algorithm 1 Real-time wide-to-PTZ acquisition

Init f_θ (wide), $\mathcal{T}$ (tracker), g (PTZ), $\mathcal{D}$ (data)
for each wide frame I_t^w **do**
  $D_t \leftarrow f_\theta(I_t^w)$; $\mathcal{T} \leftarrow \text{UpdateTracks}(\mathcal{T}, D_t)$
  Compute $U(o)$ for tracks (conf, size, novelty)
  Select $o^\star \leftarrow \arg\max U(o)$ s.t. availability
  if $U(o^\star) > \tau_{\text{trigger}}$ **then**
    Predict $\hat{d}_w(t+\Delta)$; map to (θ, ϕ, z)
    Command PTZ: $g(\theta, \phi, z)$; get $I_{t'}^p$, $\tau_{t'}$
    Run PTZ detector: $D_{t'}^p \leftarrow f_{\theta_p}(I_{t'}^p)$
    if Quality OK AND $\max s^p > \tau_{\text{confirm}}$ **then**
      $b^{w \leftarrow p} \leftarrow \text{Project}(b^p, \tau_{t'})$
      Commit $(I_t^w, b^{w \leftarrow p}, c)$ to $\mathcal{D}$
    end if
  end if
end for
Periodically retrain f_θ on $\mathcal{D}$

or the PTZ detector disagrees with the wide detector class in a way indicative of confusion (e.g., airplane vs bird). Each gate is implemented as a deterministic predicate so that dataset inclusion is reproducible and auditable.

C. Dataset Schema

Each committed example stores the wide frame (or short clip), the transferred label, and acquisition metadata. Table 1 defines a minimal schema sufficient for training and analysis.

VIII. Airplanes-in-the-Sky Test Case

A. Operating Conditions and Target Characteristics

Airplanes introduce systematic small-object issues. In a wide-angle view, aircraft can be extremely small with strong scale variation as they traverse the sky and change apparent altitude and distance (Fig. 1). Visual appearance changes with viewing angle, contrails, lighting, haze, and compression artifacts, matching known difficulty factors for small objects (low detail, interference, background ambiguity).

Flying-object detection has been studied with YOLOv8 as a practical real-time architecture choice, and YOLO-family surveys emphasize why these models are frequently deployed in resource-constrained real-time settings.

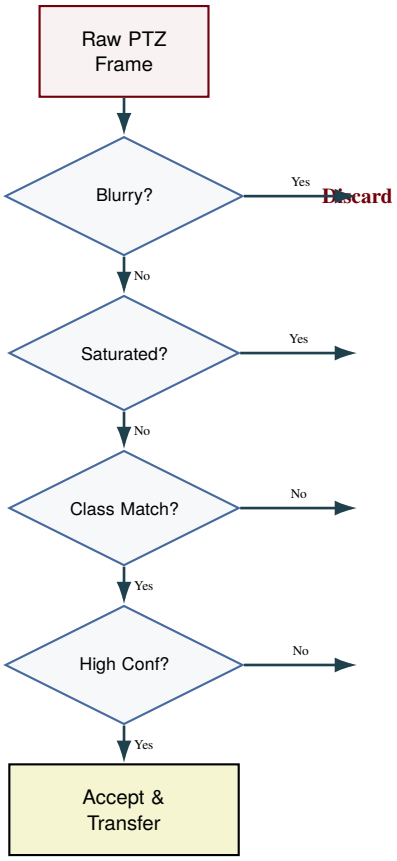

Fig. 7 Quality Gate Logic. To prevent dataset contamination (drift), PTZ frames must pass a series of deterministic checks for image quality (blur, saturation) and content verification before being used to generate labels.

B. Class Taxonomy and Negatives

For this test case, the primary class is *airplane*. Hard negatives arise from birds, insects near the lens, distant drones, clouds with sharp edges, sensor noise, and specular highlights (Fig. 8). The PTZ confirmation step naturally collects informative negatives: wide proposals that are rejected by PTZ as non-airplane can be stored as hard negatives with contextual metadata.

IX. Evaluation Protocol

This paper describes a system design and an evaluation plan intended to be executed on a real deployment. The core evaluation objective is to measure whether PTZ-assisted acquisition yields higher-quality labels and better downstream small-object performance than passive wide-only collection.

A. Baselines

The following baselines support attribution of improvements to PTZ zoom verification rather than to data volume alone:

A wide-only baseline that logs candidate detections from the wide camera without PTZ confirmation;

a PTZ patrol baseline that performs a scripted scan independent of wide detections;

a manual-zoom oracle baseline on a small audited subset, used only to estimate upper bounds and error modes.

Table 1 Minimal dataset record schema for each accepted acquisition.

Field	Description
timestamp_w	Wide-frame timestamp (monotonic)
image_w	Wide image (or video clip key)
bbox_w	Label box in wide coordinates
class	Target class (airplane for the test case)
score_w	Wide detector confidence at time t
timestamp_p	PTZ-frame timestamp used for confirmation
telemetry_p	PTZ pan/tilt/zoom, focus, exposure
bbox_p, score_p	PTZ detector box and confidence
quality_metrics	Blur, saturation, occlusion flags
site_meta	Location, camera orientation, weather

B. Metrics

Label quality is assessed on an audited subset using human review or higher-resolution reference footage. Primary metrics include (i) label precision and recall for accepted samples, (ii) bounding-box IoU between transferred labels and audited labels, (iii) confidence uplift $\Delta s = s^p - s^w$ for confirmed samples (Fig. 9), and (iv) downstream detector performance (e.g., small-object mAP on wide-angle imagery) after training on the constructed dataset.

Because PTZ tracking introduces dynamics and latency, system metrics include slew-to-lock time, dwell time per target, fraction of triggers that successfully reacquire the target, and PTZ availability contention (fraction of time PTZ is busy). PTZ tracking literature suggests these measures are essential to understanding real-world performance beyond static detection accuracy.

C. Ablation Factors

Ablations should isolate contributions from (i) trigger thresholds and hysteresis, (ii) motion prediction vs reactive centering, (iii) zoom scheduling, (iv) label transfer method (direction-only vs depth-aware if depth is available), and (v) pseudo-label admission thresholds. Adaptive pseudo-label thresholding methods provide a useful reference point for designing these ablations.

X. Implementation Considerations

A. Real-Time Constraints

The wide-angle detector must run at frame rate to maintain coverage. YOLO-family models are commonly selected for this regime; Ultralytics documentation notes YOLOv8 release and positioning as a speed–accuracy option for detection tasks, and survey work summarizes the evolution of YOLO variants and real-time deployment considerations.

The PTZ control loop must tolerate latency in (i) wide detection, (ii) command transmission, (iii) mechanical motion, and (iv) PTZ video encoding/decoding. A practical design uses a bounded queue for PTZ commands, drops stale commands, and prioritizes keeping the target near the PTZ image center over maximizing instantaneous zoom.

B. Telemetry Synchronization

Accurate label transfer requires time alignment between wide frames, PTZ frames, and PTZ telemetry. The system should log monotonic timestamps at capture and at inference, and it should record the PTZ telemetry state at the exact time a PTZ frame is captured (or as close as the interface permits). If telemetry is sampled asynchronously, interpolation to the frame time reduces geometric error.

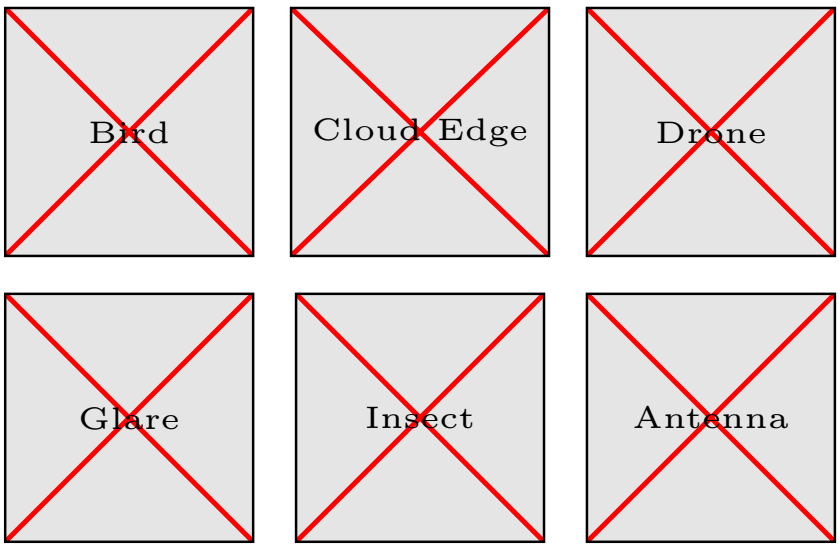

Fig. 8 Hard negative examples. Common false positives in the wide camera include birds, cloud edges, drones, sun glare, insects near the lens, and distant antennas. PTZ verification rejects these as non-airplane.

C. Safety and Privacy

The airplane test case naturally focuses on sky regions; nonetheless, deployment should constrain PTZ tilt limits and define privacy-preserving regions to avoid capturing ground-level imagery. These constraints can be implemented at the control layer by rejecting commands that enter prohibited zones.

XI. Discussion

A. When PTZ Verification Helps Most

PTZ verification is most valuable when the wide detector frequently encounters ambiguous, small candidates whose pixel support is insufficient for confident classification. In such regimes, zoom provides additional evidence without changing the wide camera that will ultimately run the detector. This can be especially beneficial when the deployment environment differs from training data and when pseudo-labeling would otherwise be unreliable.

The approach is aligned with modern PTZ-assisted perception systems that explicitly integrate deep detection with PTZ imaging to improve small-target observability.

B. Limitations

The system is constrained by single-PTZ availability and cannot zoom multiple targets simultaneously. Fast-moving targets may leave the PTZ field of view during slew, particularly when latency is high or the wide tracker is unstable. Atmospheric turbulence and haze can reduce the effective benefit of zoom. Direction-only label transfer assumes small parallax; large baselines or nearer targets require depth-aware mapping.

C. Extensions

Multi-PTZ configurations can reduce contention and increase throughput. Joint scheduling across targets can be formulated as a knapsack-like selection over predicted utilities. More advanced multi-view techniques may reduce reliance on explicit calibration in some settings, although the sky-target scenario is favorable for calibration-based geometry due to strong distance scale separation.

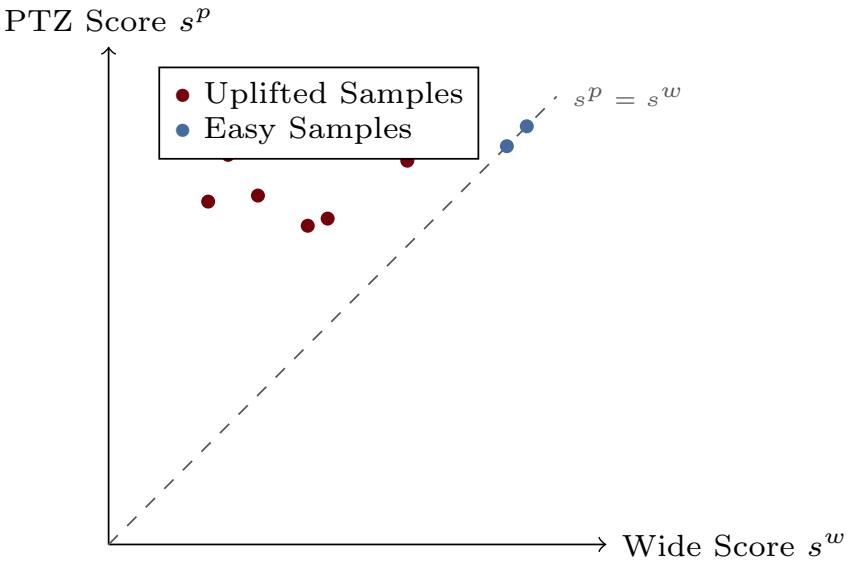

Fig. 9 Confidence uplift visualization. Points above the diagonal $s^p = s^w$ indicate samples where PTZ verification increased detection confidence. Low-confidence wide detections (left region) can achieve high PTZ confidence, validating the zoom-based verification approach.

XII. Conclusion

This paper presented a dual-camera active acquisition method for fully automated small-object dataset construction using a fixed wide-angle camera with real-time detection and a PTZ camera for zoom-based verification. For airplanes in the sky, the PTZ view provides higher-resolution evidence that can be transferred back to wide-angle frames to produce higher-quality labels, hard negatives, and metadata. We detailed calibration and control formulations, label transfer procedures, and drift-control gates grounded in pseudo-labeling and PTZ tracking insights from the literature. The proposed evaluation protocol measures label quality, confidence uplift, and downstream small-object performance, enabling rigorous assessment of whether sensor-driven "auto-labeling" can replace or substantially reduce manual annotation in small-object regimes.

References

[1] M. Nikouei et al., "Small Object Detection: A Comprehensive Survey on Challenges, Techniques, and Real-World Applications," *Intelligent Systems with Applications*, vol. 25, 2025. doi: 10.1016/j.iswa.2025.200561

[2] B. Mahaur, N. Singh, and K. K. Mishra, "Road object detection: a comparative study of deep learning-based algorithms," *Multimedia Tools and Applications*, vol. 81, no. 10, pp. 14247-14282, 2022. doi: 10.1007/s11042-022-12447-5

[3] A. Rozantsev, V. Lepetit and P. Fua, "Detecting Flying Objects Using a Single Moving Camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 879-892, 1 May 2017. doi: 10.1109/TPAMI.2016.2564408

[4] G. Chen, H. Pu, W. Luo, and L. Zhang, "A Survey of the Four Pillars for Small Object Detection: Multiscale Representation, Contextual Information, Super-Resolution, and Region Proposal," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 2, pp. 936-953, Feb. 2022. doi: 10.1109/TSMC.2020.3005231

[5] X. Zhang, Q. Chen, R. Ng, and V. Koltun, "Zoom to Learn, Learn to Zoom," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3762-3770, 2019. doi: 10.1109/CVPR.2019.00388

[6] "Real-Time Flying Object Detection with YOLOv8," *arXiv preprint arXiv:2305.09972*, 2023. doi: 10.48550/arXiv.2305.09972

[7] "Evaluation of trackers for Pan-Tilt-Zoom Scenarios," *arXiv preprint arXiv:1711.04260*, 2017. doi: 10.48550/arXiv.1711.04260

[8] "Reproducible Evaluation of Pan-Tilt-Zoom Tracking," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2015. doi: 10.1109/ICIP.2015.7351162

[9] "Active Visual Perception Enhancement Method Based on Deep Reinforcement Learning," *Electronics*, vol. 13, no. 9, p. 1654, 2024. doi: 10.3390/electronics13091654

[10] "Anomalous object detection by active search with PTZ cameras," *Expert Systems with Applications*, vol. 184, p. 115150, 2021. doi: 10.1016/j.eswa.2021.115150

[11] "VIGIA-E: Density-Aware Patch Selection for Efficient Video Surveillance with PTZ Cameras," in *Proceedings of the 25th International Conference on Computer Analysis of Images and Patterns (CAIP)*, 2025. doi: 10.1007/978-3-032-04968-1_23

[12] "Adaptive Self-Training for Object Detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2023. doi: 10.1109/ICCVW60793.2023.00102

[13] "Improving Object Detection Accuracy with Self-Training Based on Bi-Directional Pseudo Label Recovery," *Electronics*, vol. 13, no. 12, p. 2230, 2024. doi: 10.3390/electronics13122230

[14] "Ten Years of Active Learning Techniques and Object Detection: A Comprehensive Survey," *Applied Sciences*, vol. 13, no. 10, p. 6181, 2023. doi: 10.3390/app13169110

[15] "Plug and Play Active Learning for Object Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. doi: 10.1109/CVPR52733.2024.01684

[16] "Pelco Esprit Compact PTZ Technical Specifications," Pelco by Motorola Solutions, 2024. [Online]. Available: https://www.pelco.com/products/cameras/esprit-compact

[17] "Understanding Latency in IP Video Systems," PTZOptics White Paper, 2023. [Online]. Available: https://ptzoptics.com/latency/

[18] "NVIDIA Jetson Orin Nano AI Performance Benchmarks," NVIDIA Developer Blog, 2024. [Online]. Available: https://developer.nvidia.com/embedded/jetson-benchmarks