Advances in Computer, Signals and Systems (2026)
Clausius Scientific Press, Canada
DOI: 10.23977/acss.2026.100209
ISSN 2371-8838 Vol. 10 Num. 2

Static Target Snapshot Using Monocular PTZ Cameras: A Brief Review of Hardware Challenges and Observation Strategies

Minghao Zhang[a], Yang Yang*

Laboratory of Pattern Recognition and Artificial Intelligence, School of Information Science and Technology, Yunnan Normal University, Kunming, 650500, China
*[a]493427205@qq.com, *yyang_ynu@163.com*
**Corresponding author*

Keywords: Monocular PTZ camera; target snapshot; static target observation; pan–tilt–zoom control; active vision; intelligent surveillance; real-device challenges

Abstract: Monocular pan–tilt–zoom (PTZ) cameras are widely used in intelligent surveillance, long-range observation, facility inspection, and target detail acquisition because they can actively adjust both viewing direction and imaging scale. Compared with fixed cameras, PTZ cameras provide greater flexibility for capturing useful target images in large-scale scenes. However, static target snapshot with PTZ cameras is not a simple target-centering problem. A useful snapshot should keep the target complete, provide sufficient image scale, and remain robust to real-device errors. This paper provides a brief review of static target snapshot tasks with monocular PTZ cameras. First, the main hardware and system challenges are summarized, including limited field of view, pan–tilt motion errors, zoom-related optical changes, control delay, video stream synchronization, and perception uncertainty. Then, static target snapshot tasks are discussed from single-target and multi-target perspectives. Static single-target snapshot focuses on acquiring a clear and complete view of one target, while static multi-target snapshot further involves target selection, view switching, capture order, zoom planning, and motion-cost reduction. Finally, several future directions are discussed, including full PTZ snapshot planning, real-device-aware modeling, snapshot-oriented evaluation, and open-source experimental platforms.

1. Introduction

With the development of intelligent surveillance, long-range observation, traffic monitoring, ecological inspection, and public safety applications, visual systems are increasingly required not only to detect targets, but also to acquire clear and useful target images at suitable moments. Fixed cameras can provide stable wide-area monitoring, but their viewing direction and imaging scale are fixed. As a result, distant or small targets often lack sufficient visual details for recognition, verification, or evidence recording.

Pan–tilt–zoom cameras provide a more flexible solution for active visual observation. A PTZ camera can change its viewing direction through pan and tilt motion, and can adjust the target imaging scale through optical zoom. Therefore, it can actively select a target region and zoom in to acquire a clearer snapshot. Recent studies have shown that PTZ cameras are useful for bridging wide-area observation and detailed target acquisition, especially in large-scale surveillance scenarios where fixed wide-view images are often insufficient for fine-grained analysis [1-2].

In this review, the static target snapshot task refers to the process in which a monocular PTZ camera actively adjusts its pan, tilt, and zoom states to capture one or several useful images of static or quasi-static targets. Here, a static target means that the target position remains unchanged, or its motion during the camera control and image acquisition interval is small enough to be ignored. Typical examples include fixed facilities, parked vehicles, stationary abnormal objects, signboards, inspection targets, and temporarily stationary animals or persons.

This task is different from ordinary object detection. Detection only estimates where the target is, whereas snapshot acquisition further considers whether the captured target image is clear, complete, and informative. It is also different from long-term tracking, because snapshot acquisition emphasizes obtaining a high-quality image within a limited observation process, rather than continuously following a moving target over time.

According to the number of targets, static snapshot tasks can be divided into single-target snapshot and multi-target snapshot. In a single-target task, the PTZ camera focuses on one target and attempts to obtain a clear and complete view of it. In a multi-target task, several targets need to be captured in the same scene. Since a monocular PTZ camera observes only one limited field of view at a time, especially under high zoom, the system must consider target priority, view switching order, zoom level, and camera motion cost. Recent work on rapid multi-target capture using a single PTZ camera shows that such tasks should be treated as both a target acquisition problem and a viewpoint scheduling problem [1].

However, PTZ target snapshot is strongly affected by real hardware limitations. Real PTZ cameras are usually controlled through SDKs, ONVIF protocols, or private interfaces, and their internal mechanical and optical models are often unavailable. Pan–tilt motion may suffer from angle quantization, backlash, repeated positioning error, and control delay. Zoom changes not only target scale, but also field of view, camera intrinsics, focus state, and sensitivity to control error. Calibration and auto-calibration studies also show that estimating camera parameters for arbitrary PTZ viewpoints remains an important problem in real applications [3-4]. In addition, industrial patents have described practical PTZ adjustment and 3D positioning schemes based on image coordinates, target regions, and camera parameters, indicating that geometry-based PTZ control remains important in real surveillance systems [7-8].

Learning-based PTZ control has also attracted attention. Some studies use reinforcement learning to enhance low-confidence targets or support active visual perception with PTZ cameras [2], [5]. Earlier DQN-based PTZ control research also treated PTZ camera control as a sequential decision-making problem, where target position and scale information are used to select pan, tilt, and zoom actions [6]. These works show the potential of data-driven active observation, but they still face challenges such as training cost, device dependency, and limited interpretability.

Based on the above background, this paper provides a brief review of static target snapshot tasks with monocular PTZ cameras. Different from general PTZ control discussions, this review does not reduce PTZ cameras to pan–tilt devices, nor does it treat snapshot acquisition as simple target centering. Instead, it first analyzes hardware and system challenges, and then discusses static single-target and static multi-target snapshot tasks.

2. Hardware and System Challenges of Monocular PTZ Cameras

Before discussing static snapshot tasks, it is necessary to clarify the hardware and system challenges of monocular PTZ cameras. A PTZ camera is not only an imaging sensor, but also an active mechanical-optical system. Its observation result is jointly affected by view direction, zoom state, mechanical execution, video delay, and visual perception. Therefore, PTZ target snapshot is a coupled perception-control-imaging problem.

2.1 Limited Field of View and Single-View Observation

A monocular PTZ camera observes only one view at a time. When it uses a wide field of view, it can cover a larger scene, but distant targets may be too small. When it zooms in, the target becomes larger, but the visible area becomes narrower. This creates a fundamental trade-off between global awareness and local detail.

This limitation is important for both single-target and multi-target snapshot tasks. For a static single target, a narrow view may be acceptable if the target location is stable. However, for static multi-target tasks, zooming in on one target may cause other targets to leave the view. Therefore, the best snapshot action is not always the one with the largest zoom. It should balance target detail, view safety, and observation context [1].

2.2 Pan–Tilt Motion Errors

In ideal geometric models, pan and tilt are often treated as accurate rotations around a fixed optical center. However, real PTZ devices do not strictly satisfy this assumption. Their motion may be affected by finite angular resolution, mechanical backlash, repeated positioning errors, motor response delay, and limited accuracy of returned PT states.

These errors become more serious under high zoom. A small angular error may correspond to a large image displacement when the field of view is narrow. If f_u and f_v denote the horizontal and vertical focal lengths in pixel units, small pan and tilt errors may approximately produce image displacements:

$$\delta u \approx f_u \tan(\delta p),$$

$$\delta v \approx f_v \tan(\delta t).$$

As zoom increases, the effective pixel focal length usually increases. Thus, the same angular error may cause a larger pixel shift. This explains why close-up snapshot is more sensitive to pan–tilt control error than wide-view monitoring.

2.3 Zoom-Related Optical Changes

Zoom is a key factor that distinguishes PTZ cameras from ordinary pan–tilt cameras. In target snapshot tasks, zoom determines the target imaging scale. However, zoom is not merely an image magnification operation. It changes the field of view, focal length, image scale, focus behavior, and tolerance to error.

For a simplified camera model, the relationship between focal length and field of view can be written as

$$FOV_i(z) = 2\arctan\left(\frac{l_i}{2f_i(z)}\right), i \in \{h, v\},$$

where $FOV_i(z)$ is the horizontal or vertical field of view at zoom state z, $l_i$is the effective imaging-plane size, and $f_i(z)$ is the effective focal length. This relation shows that when the focal length increases, the field of view decreases.

In real PTZ cameras, the zoom command may not be linearly related to the true optical magnification. Camera intrinsics may also change with zoom. In addition, the effective zoom center may deviate from the image center, causing the target to shift during zooming. Therefore, zoom control must balance detail enhancement and target retention. Recent PTZ calibration and auto-calibration research also shows that robust parameter estimation for arbitrary PTZ viewpoints remains important for real applications [3-4].

2.4 Control Delay and Video Stream Synchronization

PTZ cameras do not change their states instantly. After a control command is sent, the camera needs time for mechanical movement, stabilization, autofocus, exposure adjustment, and frame update. Meanwhile, the video stream may contain buffered frames. Therefore, the received image may not correspond exactly to the latest physical PTZ state.

For static targets, delay is usually less serious than in dynamic tracking tasks, but it still affects snapshot reliability and evaluation. If the captured frame, PTZ state, and detection result are not synchronized, the measured snapshot quality may not reflect the actual control result. A practical system should wait for camera stabilization, flush old frames when necessary, and record PTZ states and images consistently.

2.5 Perception Errors

PTZ snapshot depends on visual perception results, such as target centers, bounding boxes, confidence scores, or segmentation masks. Detection errors may come from small target size, low contrast, partial occlusion, background clutter, or bounding-box instability. These errors can propagate to the final PTZ command.

Zoom also changes the behavior of perception algorithms. A target may become easier to recognize after zooming in, but excessive zoom may crop the target and cause detection failure. Therefore, perception and control influence each other. Recent PTZ visual enhancement studies have considered this perception-control interaction by using PTZ actions to improve low-confidence observations [2], [5]. Earlier DQN-based PTZ control research also used target position and scale information as the state input for selecting pan, tilt, and zoom actions, showing that learning-based methods can be applied to PTZ control decisions [6].

In summary, monocular PTZ snapshot is affected by limited field of view, pan–tilt motion errors, zoom-related optical changes, system delay, stream synchronization, and perception uncertainty. These factors should be considered before analyzing static target snapshot tasks.

3. Static Target Snapshot Tasks

Static target snapshot is a basic form of monocular PTZ active observation. In this task, the target position is assumed to remain unchanged, or its motion during camera control and image acquisition is small enough to be ignored. The goal is to acquire a usable target image for recognition, inspection, or evidence recording. Therefore, the target should remain inside the field of view, have sufficient image scale, and avoid serious cropping or poor composition.

3.1 Static Single-Target Snapshot

Static single-target snapshot is the simplest case. The system focuses on one detected or manually specified target. In the initial wide-view image, the target may be far from the image center or may occupy only a small region. The PTZ camera then adjusts its viewing direction and zoom state to capture a clearer image.

In this task, pan and tilt mainly determine the target position in the final view, while zoom determines the target scale. A larger zoom can provide more details, but it also narrows the field of view and reduces tolerance to control errors. Therefore, the objective is not simply to maximize zoom, but to obtain a clear and complete target image.

Real systems are affected by camera calibration, zoom variation, and device repeatability. If the relationship among image coordinates, pan–tilt motion, and zoom state is inaccurate, the final target position may deviate from the expected view. Existing PTZ calibration, auto-calibration, and visual enhancement studies suggest that static snapshot should be treated as an active image acquisition process, rather than a passive detection result [2]–[5]. Industrial PTZ adjustment and positioning patents also reflect the practical need to convert image-region or coordinate information into camera control results in surveillance systems [7-8].

3.2 Static Multi-Target Snapshot

Static multi-target snapshot is more complex. In this task, multiple targets exist in the scene, and the PTZ camera is expected to acquire useful images of several or all of them. Since a monocular PTZ camera observes only one limited field of view at a time, especially under high zoom, it cannot always capture all targets with sufficient detail in a single shot.

Therefore, static multi-target snapshot is closer to an observation scheduling problem. The system needs to decide which target should be captured first, whether nearby targets can be captured together, how much zoom should be used, and how to reduce unnecessary pan–tilt–zoom movement. If targets are close to each other, a moderate zoom may cover several targets in one view. If they are far apart, the camera may need to visit them sequentially.

The main difference between single-target and multi-target snapshot lies in view planning. For one target, the system only optimizes the observation of that target. For multiple targets, the system must consider target priority, camera motion cost, total acquisition time, and possible loss of scene context. Recent work on rapid multi-target capture with a single conventional PTZ camera directly addresses this issue by considering target capture quality and realistic PTZ motion cost [1].

3.3 Summary of Static Snapshot Tasks

Static target snapshot still faces several difficulties. First, target scale and field of view must be balanced. Conservative zoom improves safety but may reduce target detail, while aggressive zoom improves detail but increases cropping risk. Second, final snapshot quality depends on real-device behavior, including pan–tilt error, zoom-dependent field-of-view variation, and imperfect calibration. Third, multi-target static snapshot introduces scheduling difficulty. A strategy that is good for one target may not be optimal for all targets. Finally, perception uncertainty still exists even for static targets, because inaccurate detection boxes may lead to inaccurate PTZ commands.

In summary, static target snapshot is not merely target centering. Static single-target snapshot focuses on acquiring a clear and complete view of one target, while static multi-target snapshot further introduces view scheduling and motion-cost constraints.

4. Discussion and Future Directions

Static target snapshot with monocular PTZ cameras is a practical but still challenging active observation problem. Although the target is assumed to be static or quasi-static, the system still needs to jointly consider target visibility, imaging scale, field-of-view safety, real-device errors, and multi-target scheduling. Therefore, future studies should not treat static snapshot as a simple target-centering problem. Instead, it should be studied as a complete PTZ-based image acquisition process.

4.1 From Pan–Tilt Control to Full PTZ Snapshot

Many simple PTZ systems mainly focus on pan–tilt adjustment, where the objective is to move the target toward a desired image position. However, for snapshot tasks, zoom is equally important. Pan and tilt determine the target location in the image, while zoom determines the target scale and detail level. A useful snapshot should not only place the target in a proper position, but also provide enough image details without causing target cropping.

Therefore, future methods should move from PT-only control to full PTZ snapshot planning. The system should jointly consider where the target should appear, how large the target should be, and how much safety margin should be preserved. This is especially important under high zoom, where small positioning errors may lead to large image displacement.

4.2 From Ideal Geometry to Real-Device-Aware Modeling

Ideal geometric models are useful for explaining the relationship between image coordinates and camera motion. However, real PTZ cameras usually have non-ideal behaviors, including pan–tilt quantization, backlash, repeated positioning error, zoom-dependent field-of-view changes, and calibration uncertainty. These factors may cause mismatch between the planned view and the captured snapshot.

Future snapshot methods should explicitly consider real-device characteristics. One possible direction is to combine interpretable geometric models with lightweight calibration or data-driven correction. Recent PTZ calibration and auto-calibration studies have shown that robust parameter estimation for arbitrary PTZ viewpoints remains important for real applications [3-4]. Related industrial patents also reflect the practical need to convert image-region information into PTZ adjustment or 3D positioning results under real surveillance-camera settings [7-8]. Such calibration or compensation methods can improve the reliability of static target snapshot, especially under high zoom.

4.3 From Single-Target Snapshot to Multi-Target Scheduling

Static single-target snapshot mainly focuses on acquiring a clear and complete view of one target. However, many real applications involve multiple static targets, such as facility inspection, wide-area monitoring, and scene patrol. In these cases, the problem becomes more than target acquisition. It also involves target selection, view switching, capture order, zoom planning, and motion-cost reduction.

For static multi-target snapshot, the system should determine whether targets can be captured together or should be visited separately. It should also balance snapshot quality and total acquisition time. Recent work on rapid multi-target capture with a single conventional PTZ camera has shown that realistic PTZ motion cost and capture quality should be jointly considered [1]. This indicates that future multi-target snapshot systems should be designed as scheduling-aware active observation systems.

4.4 From Detection Output to Snapshot Quality Evaluation

Object detection results are often used as the input of PTZ snapshot systems. However, successful detection does not necessarily mean successful snapshot. A target may be detected in the initial image, but the final snapshot may still suffer from small target scale, partial cropping, poor focus, or large position deviation. Therefore, snapshot quality should be evaluated separately from detection accuracy.

Future evaluation protocols should include target completeness, image scale, boundary safety, final target position, and acquisition time. For multi-target tasks, additional metrics such as total capture time, number of successfully captured targets, and average motion cost should also be considered. Such evaluation criteria would make PTZ snapshot studies more consistent with real application requirements.

4.5 Open-Source Platforms and Reproducible Experiments

Another important direction is the development of open-source PTZ experimental platforms. Real PTZ experiments involve camera connection, PTZ state reading, image acquisition, command execution, frame buffering, and result logging. Without unified tools and protocols, it is difficult to compare different methods fairly.

Open-source platforms can reduce the difficulty of PTZ research and improve reproducibility. They can provide standard interfaces for device control, data recording, snapshot evaluation, and benchmark construction. For static snapshot tasks, such platforms should at least record the initial image, target annotation, initial and final PTZ states, final snapshot image, zoom value, and snapshot quality indicators. Recent PTZ visual enhancement and active perception studies also show that real-device experiments are important for evaluating PTZ-based observation methods [2], [5]. Therefore, future research should pay more attention to public tools, datasets, and evaluation protocols for static target snapshot tasks.

4.6 Summary

Overall, static target snapshot with monocular PTZ cameras should be understood as an active image acquisition problem. For single-target tasks, the key issue is to obtain a clear, complete, and stable target view. For multi-target tasks, the key issue further extends to view scheduling and motion-cost optimization. Future research may focus on full PTZ planning, real-device-aware modeling, snapshot-oriented evaluation, and open-source experimental platforms. These directions can help PTZ cameras provide more reliable target images for intelligent surveillance, inspection, and long-range observation.

5. Conclusion

This paper briefly reviews static target snapshot tasks with monocular PTZ cameras. Different from fixed cameras, PTZ cameras can actively adjust their viewing direction and imaging scale through pan, tilt, and zoom control. This capability makes them useful for long-range observation, intelligent surveillance, facility inspection, and target detail acquisition. However, static target snapshot is not a simple target-centering task. A useful snapshot should preserve target completeness, provide sufficient image scale, and remain robust to real-device errors.

This review first analyzed the hardware and system challenges of monocular PTZ cameras, including limited field of view, pan–tilt motion errors, zoom-related optical changes, control delay, video stream synchronization, and perception uncertainty. These factors show that PTZ snapshot should be regarded as a coupled perception-control-imaging problem rather than an ideal geometric control problem.

Then, static target snapshot tasks were discussed from single-target and multi-target perspectives. Static single-target snapshot focuses on acquiring a clear and complete view of one target, while static multi-target snapshot further involves target selection, view switching, capture order, zoom planning, and motion-cost reduction. Therefore, multi-target snapshot is closer to an observation scheduling problem.

Future research may focus on full PTZ snapshot planning, real-device-aware modeling, snapshot-oriented evaluation metrics, and open-source experimental platforms. These directions can help monocular PTZ cameras provide more reliable and useful target images for practical intelligent observation systems.

References

[1] F. Jiao, M. Zhang, Y. Tian, Y. Zhang, D. Ren, Y. Yang, S. Zhao, and Z. Yin, "A generalizable rapid multi-target capture framework based on a single conventional pan-tilt-zoom camera," Expert Systems with Applications, vol. 309, Art. no. 131189, 2026, doi: 10.1016/j.eswa.2026.131189.

[2] H. Fang, H. Liu, J. Wen, Z. Yang, J. Li, and Q. Han, "Automatic visual enhancement of PTZ camera based on reinforcement learning," Neurocomputing, vol. 626, Art. no. 129531, 2025, doi: 10.1016/j.neucom.2025.129531.

[3] F. Yang, H. Chen, X. Zhang, S. Wang, and L. Zhang, "SPLAC: A single-step PTZ camera linear auto-calibration method," Neurocomputing, vol. 645, Art. no. 130377, 2025, doi: 10.1016/j.neucom.2025.130377.

[4] J. Guo, L. Fan, B. Wu, J. Gu, S. Cao, and J. Ye, "PTZ-Calib: Robust pan-tilt-zoom camera calibration," in Proc. IEEE International Conference on Robotics and Automation (ICRA), 2025, pp. 1907–1913, doi: 10.1109/ICRA55743.2025.11128129.

[5] Z. Yang, H. Fang, H. Liu, J. Li, Y. Jiang, and M. Zhu, "Active visual perception enhancement method based on deep reinforcement learning," Electronics, vol. 13, no. 9, Art. no. 1654, 2024, doi: 10.3390/electronics13091654.

[6] D. Kim, K. Kim, and S. Park, "Automatic PTZ camera control based on Deep-Q Network in video surveillance system," in Proc. International Conference on Electronics, Information, and Communication (ICEIC), Auckland, New Zealand, 2019, pp. 1–3, doi: 10.23919/ELINFOCOM.2019.8706453.

[7] Huawei Technologies Co., Ltd., "Adjustment method and apparatus for pan-tilt camera," Chinese Patent CN108874823A, Sept. 25, 2018. [in Chinese].

[8] Wuhan FiberHome Digital Technology Co., Ltd., "3D positioning method and device based on PTZ surveillance camera," Chinese Patent CN103905792A, Jul. 2, 2014. [in Chinese].