Falcon: A Wide-and-deep Onboard Active Vision System

Masahiro Hirano[1] and Yuji Yamakawa[1]

***Abstract*— The tradeoff between the field-of-view and resolution of conventional onboard vision systems primarily results from their fixed optical components. We propose a novel active vision system, Falcon, as an optimal solution. This system comprises an electric zoom lens connected to a high-speed camera with a pair of galvanometer mirrors, enabling high-resolution imaging of a moving object across a wide range, from near to far. To ensure accurate calibration of the Falcon system, we introduce a novel mapping-based calibration method using external cameras. We also present a robust and lightweight visual feedback method that utilizes this mapping-based calibration for effective object tracking. The effectiveness of the Falcon system is verified by constructing a prototype and conducting tracking experiments in an indoor setting, which demonstrated the superior performance of our method. Additionally, we successfully achieved continuous and high-resolution imaging of a curved mirror on public roads while the vehicle was moving.**

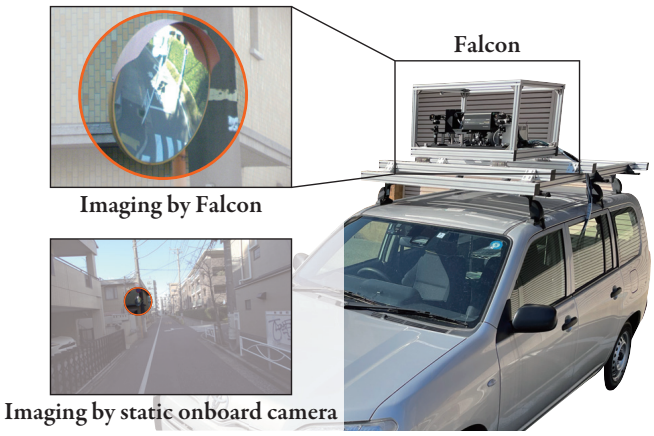

Fig. 1. Our proposed system enables high-resolution imagery of a moving object in a wide range from near to far.

I. INTRODUCTION

Sensing the surrounding environment is a fundamental function of robots, including autonomous vehicles. While recent advances in electronics have facilitated the development of various sensors such as LiDAR and millimeter radar, vision sensors play an indispensable role due to their high capability in environmental recognition, which is widely appreciated in many vehicle applications.

Conventional onboard vision sensors observe the surrounding scene uniformly with a fixed optical system, which inherently involves a tradeoff between the measurement area and resolution. However, there is a discrepancy in the importance of different parts of the surrounding environment. This suggests that by optically adjusting the measurement area based on the context of the surrounding environment, areas crucial for driving can be observed with high resolution, contributing to safer and more efficient autonomous driving. For instance, by focusing on a curved mirror, the system can observe a blind area around the corner in detail, or by magnifying the traffic light, the recognition rate can be improved.

Active optical systems have emerged as promising solutions for advanced environment-sensing applications. One such example is adaptive LiDAR, which dynamically adjusts sensing points based on a contextual understanding of a scene to enhance the accuracy of depth measurements [1], [2]. Similarly, active systems that enable dynamic control of the viewing direction have been proposed for cameras. However, the use of an electronic platform to move the entire camera system, including the housing and lens, has resulted in issues with large inertia and slow movement. To address this challenge, studies have explored the utilization of galvanometer mirrors or MEMS mirrors, which possess lower inertia, in various fields such as augmented reality [3] and image-based aerodynamic measurements [4]. These mirrors have also demonstrated potential for incorporation in onboard vision systems [5]. However, conventional active vision systems employing fixed-focus telephoto lenses have limitations in capturing distant areas in high definition due to a narrow depth of field. Furthermore, the magnification rate remains constant regardless of the distance to the object, which is suboptimal for onboard vision systems that need to observe the surrounding environment from near to far.

In this study, we introduce a novel active onboard vision system called Falcon, which overcomes the limitations of conventional systems by utilizing a 2-axis galvanometer mirror and variable optics with adjustable zoom and focal position. This advancement enables continuous observation of a wide and distant area at high resolution, contributing to safer and more efficient autonomous driving. To ensure precise and reliable performance, we developed a tracking mechanism that enables Falcon to place the object at the center of the field of view, even when the object's relative position changes rapidly due to vehicle movement or vibrations.

*This work was not supported by any organization.

[1]Masahiro Hirano and Yuji Yamakawa are with Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan. mhirano@iis.u-tokyo.ac.jp

We propose a novel mapping-based calibration method that guarantees the accuracy and precision of Falcon's performance. Additionally, we present a visual feedback method that utilizes the calibration data to optimize object tracking.

To evaluate the effectiveness of our proposed visual feedback method, we developed a prototype and successfully demonstrated the continuous tracking of a curved mirror as an example of an area in the surrounding environment while the ego vehicle was in motion. Overall, our study represents a significant advancement in the development of active onboard vision systems for autonomous vehicles, with potential applications across various industries and settings.

The main contributions of our study can be summarized as follows:

- We propose Falcon, an active onboard vision system that enables close observation of objects in the wide-and-deep surrounding environment.
- We propose an accurate mapping-based calibration method for active vision systems, along with a robust and lightweight visual feedback method for tracking using the mapping-based calibration.
- We validate the proposed system with a prototype through performance evaluation and demonstrate curved mirror tracking as an application in intelligent transportation systems (ITS).

II. RELATED WORK

Active vision systems incorporating small mirrors have seen significant advancements in recent years. For instance, Li *et al.* proposed a system utilizing a dual-axis galvanomirror-based imaging system [6]. However, this system operates independently of external cameras, requiring an additional calibration process that could introduce errors. Our calibration method inherently integrates external cameras, thereby mitigating such errors. Zhang *et al.* designed a mechanically actuated parfocal zoom lens with a dual-axis galvano mirror, demonstrating good performance by focusing on predefined static points at varying distances [7]. This mechanical design can seamlessly work with our proposed system to enable wide-and-deep object tracking.

Such active vision systems are particularly suited for object tracking. Okumura *et al.* demonstrated ping-pong ball tracking by controlling a pair of galvanometer mirrors at high speed [8]. While this simple method is ideal for tracking simple high-speed moving objects, tracking various objects in natural environments necessitates machine learning-based recognition. To this end, Hu *et al.* developed a deep learning-based multi-object observation system, offering superior performance in complex environments [9]. Galvanometer mirror-based active vision systems have been improved using pupil shift lenses, which align the optical center with the mirror center, thereby enhancing the field of view [8]. However, such optics are not required for our system, as we do not need imagery from small-focal-length lenses. Another advancement in this domain involves using a three-mirror system, enabling 360° imagery [10].

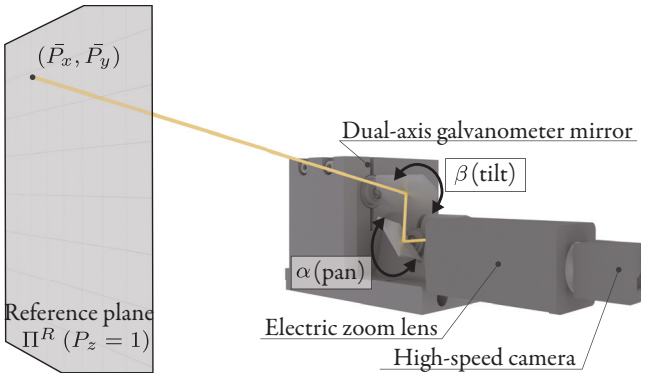

Fig. 2. Schematic drawing of Falcon. The internal camera observes the foreground through a variable zoom lens and a pair of fast-tilting mirrors.

Recently, active vision systems based on MEMS mirrors have emerged as a promising approach due to their compactness, making them ideal for onboard vision systems. Tilmon *et al.* proposed a system using a MEMS mirror to simultaneously observe multiple static objects by tilting the mirror [11]. They also demonstrated how such systems could enhance the accuracy of depth estimation in the surrounding environment [5].

In this study, we introduce a novel active vision system designed for continuous, close observation of a wide and deep surrounding environment. This system combines a pair of fast-tilting mirrors with a variable zoom lens. We also propose an accurate calibration method and a robust, lightweight visual feedback method for object tracking to operate the system.

III. FALCON: A WIDE-AND-DEEP ACTIVE VISION SYSTEM

A. Components

The Falcon system is comprised of a dual-axis galvanometer mirror and an electrically controllable zoom lens attached to a high-speed camera (internal camera). As shown in Fig. 2, we carefully aligned the optical axis of the internal camera to intersect the center of the first mirror and remain parallel to the axis of the second mirror. By changing its gaze direction through a pair of reflections from the fast-tilting mirrors, the internal camera observes the foreground scene. Furthermore, two wide-angle video rate cameras (external cameras) are mounted on the Falcon system. These cameras survey the foreground scene to identify areas requiring close observation.

B. Calibration

1) Model-based calibration: For calibration, we used an optical model originally developed for a Saccade Mirror [8], which also incorporates a pair of galvanometer mirrors. During the model-based calibration process, the

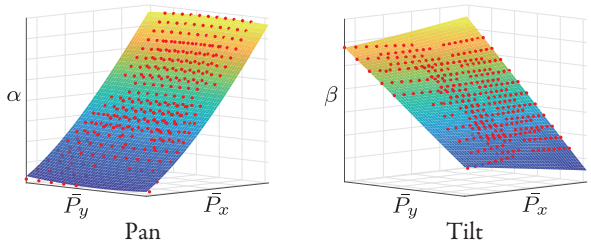

Fig. 3. An example of mappings from (P_x, P_y) to (α, β) and its fitted cubic polynomial surfaces (colored in gradation). Each red point depicts a corner point on the board.

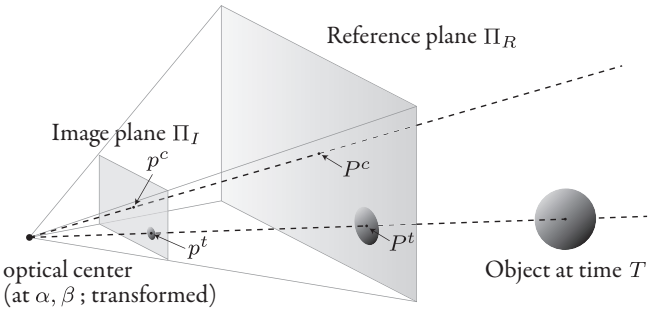

Fig. 4. Schematic illustration for mapping-based control. The internal camera's optical center is transformed according to the mirrors' coordinate system at time T, where pan and tilt angles are α and β, respectively.

zoom ratio (R_Z) remained constant. Before calibrating the Falcon, we first conducted stereo calibration of the two wide-angle cameras. For this, we placed a large calibration board in front of the system and scanned the board with the Falcon to record the mirror angles, the recognized corner point's 3D position (determined through triangulation), and its projected point on the internal camera. We then calculated the calibration parameters through optimization using the model and measurements.

These parameters allowed us to derive a model-based projection

$$p = \pi(P; R_Z), \tag{1}$$

which maps a point P on a reference plane Π_R to a point p on an image plane Π_I, where Π_R is a plane parallel to the P_x-P_y plane, that is, $P_z = 1$. However, owing to modeling errors, the accuracy of the model-based calibration was insufficient.

2) Mapping-based calibration: To enhance the calibration accuracy, we propose a method that refines the projection through an additional calibration process. Initially, we captured a pair of images of the calibration board using the wide-angle cameras, allowing us to determine the 3D positions (P_x, P_y, P_z) of the corner points on the board, placed in front of the system. For each corner point, we adjusted the mirror angles α (pan) and β (tilt) to center the corner point, using parameters from the model-based calibration. We made delicate adjustments to these angles so that the image center coincided with the corner point. This process yielded a refined mapping $M : (\bar{P}_x, \bar{P}_y) \to (\alpha, \beta)$, where $(\bar{P}_x, \bar{P}_y) = (P_x/P_z, P_y/P_z)$ is the projected corner point on the reference plane Π_R. This mapping accurately reflects the gaze direction relative to the mirror angles. By repeating this process, we obtained a collection of mappings M_i. We then approximated these mappings M_i using a pair of continuous functions μ_α and μ_β based on cubic polynomial surfaces with bisquare robust estimators, as shown in Fig. 3:

$$(\alpha, \beta) = \mu(\bar{P}_x, \bar{P}_y) = \left(\mu_\alpha(\bar{P}_x, \bar{P}_y), \mu_\beta(\bar{P}_x, \bar{P}_y)\right). \tag{2}$$

C. Tracking Strategy

To observe a target object using Falcon, we employed the strategy described below. The process involves detection and tracking the object, and there are three possible statuses: *IDLE*, *DETECTED*, and *TRACKING*.

1) IDLE: In this state, the target has not yet been detected. We capture a pair of images using the external cameras and input them into the object detection module. If the target is identified in both images, the state transitions to *DETECTED*.

2) DETECTED: Once the target is detected, we use its triangulated 3D position to control the galvanometer mirrors and zoom lens, enabling Falcon to focus on the object in a feed-forward manner. We also initiate object detection for the internal camera images. If the target is detected in the internal camera image, the state transitions to *TRACKING*. If object detection repeatedly fails, the state reverts to *IDLE*.

3) TRACKING: In this state, Falcon tracks the target using feedback control. When the target is detected in the internal camera image, we adjust the mirrors to keep the object at the center of the image using a feedback mechanism. If object detection fails repeatedly, the state reverts to *IDLE*. Object detection is performed asynchronously, and the results are retrieved when available. It is important to note that the external camera images are always used to provide depth information of the target, regardless of the tracking status, which is essential for controlling the zoom lens.

D. Control

For tracking control, we employ a combination of feed-forward and feedback controls, depending on the tracking status. When the control strategy is set to feed-forward, we determine the mirror angles α and β based on the position of the target object on the reference plane $(\bar{P}_x, \bar{P}_y)$ detected by external cameras:

$$(\alpha, \beta) = \mu(\bar{P}_x, \bar{P}_y). \tag{3}$$

Additionally, we control the zoom ratio and focus position in a feed-forward manner, utilizing the distance to the target as determined by the external cameras. In the case of a tracking status set to *TRACKING* and a control strategy set to feedback, the mirror angles are quickly adjusted to respond to the target location

detected by the internal camera. To adjust the mirror angles, a common approach in active vision systems is to use a PID controller to minimize the distance to the center. However, tuning the control parameters can lead to instability. In this study, we propose a novel mapping-based visual feedback method for active vision systems that maximizes the utilization of accurate calibration. Let $p^t \in \Pi_I$ denote the location of the target at time T, with mirror angles α and β. Similarly, let $p^c \in \Pi_I$ represent the center of the image (see Fig. 4).To capture the object at the center, we need to determine new mirror angles α' and β'. This is equivalent to obtaining the object's location on the reference plane, denoted as $P^t = \mu^{-1}(\alpha', \beta')$. However, directly obtaining P^t from p^t using μ^{-1} is not possible because μ describes the relationship between the mirror angles and a point on the reference plane, which is always mapped to the center of the image.

To address this issue, we assumed the accuracy of the model-based calibration. It is assumed that the distances between points $\mu^{-1}(\alpha', \beta')$, $\mu^{-1}(\alpha, \beta)$ are approximately equal to the distances between points $\pi^{-1}(p^t; R_Z)$, $\pi^{-1}(p^c; R_Z)$:

$$\mu^{-1}(\alpha', \beta') - \mu^{-1}(\alpha, \beta) \simeq \pi^{-1}(p^t; R_Z) - \pi^{-1}(p^c; R_Z), \quad (4)$$

where R_Z is the zoom ratio at time T, which is updated in real time. The approximate mirror angles are obtained as follows:

$$\begin{aligned}(\alpha', \beta') &= \mu(P^t) \\ &= \mu\left(P^c + (P^t - P^c)\right) \\ &= \mu\left(\mu^{-1}(\alpha, \beta) + \mu^{-1}(\alpha', \beta') - \mu^{-1}(\alpha, \beta)\right) \\ &\simeq \mu\left(\mu^{-1}(\alpha, \beta) + \pi^{-1}(p^t; R_Z) - \pi^{-1}(p^c; R_Z)\right) \quad (5)\end{aligned}$$

We approximate the inverse of μ similarly to μ, leveraging its injectivity, as depicted in Fig. 3. Given the algebraic and precomputable nature of all functions in Eq. (5), the computational cost is minimal. Moreover, unlike PID control, there's no need for user-tuned parameters, an advantage for practical use. As the acquisition frame rate increases, tracking robustness improves and the error of the target position from the center $\|p^t - p^c\|$ diminishes, enhancing the approximation accuracy in Eq. (4).

IV. SYSTEM EVALUATION

A. Overview

To evaluate our system, we built a prototype shown in Fig. 5. We assessed the mapping-based feedback control by tracking an indoor tennis ball. We compared this with feed-forward control, which adjusts mirror angles based on the ball's detected position via external cameras, and with PID-based feedback. The ball moved along a large horizontal circle while rotating on a small vertical circle.

B. Prototype Configuration

We designed a Falcon prototype using a dual-axis galvanometer mirror (Cambridge Technology model:6240H, aperture:30 mm, rated excursion:±20°). This mirror was fitted with a high-speed Basler camera (model:acA800-510uc, resolution:800 × 600 px, frame rate:100 fps) and an electrically adjustable zoom lens from Fujinon (model:FH32x15.6SR4A-CV1, focal length:15.6−500 mm, F:3.9−16), as seen in Fig. 5. Additionally, two synchronized LUCID cameras (model:TRI054S-CC, resolution:2880 × 1860 px, frame rate:10 fps) with wide-angle lenses (focal length:5 mm, F:2.8), triggered externally. The external cameras had a baseline of approximately 462 mm, with the left camera positioned 298 mm from the galvanometer mirror. The entire system was mounted on an optical breadboard (600 × 450 mm).

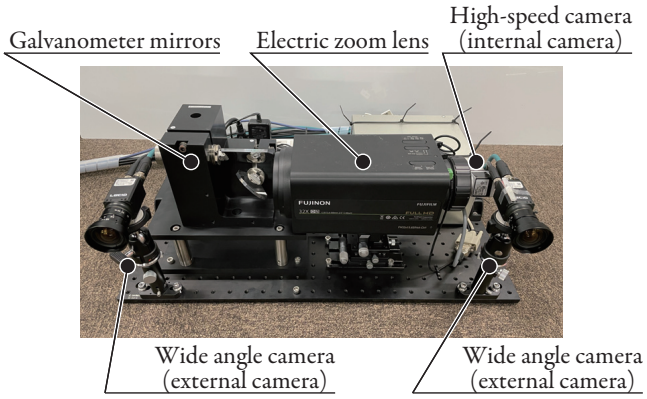

Fig. 5. A prototype of Falcon.

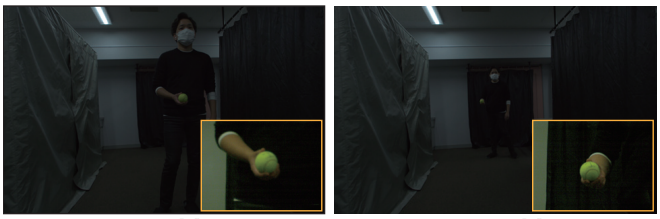

Fig. 6. Captured images by the Falcon and the external camera while tracking with the proposed mapping-based control. The image captured by the Falcon is shown as an inset of the image captured by the external camera.

The Falcon system operates on a computer (OS: Windows 10; CPU: Intel Xeon Gold 6128; GPU: NVIDIA RTX3090, Memory: 192 GB). We employed YOLOv8-s [12] with pre-trained weights for object detection, compiled with NVIDIA's TensorRT [13] to enhance inference speed.

C. Results

Fig. 6 displays images captured by the Falcon and an external camera while tracking the ball with our proposed mapping-based feedback control. The ball maintained nearly the same size in the frame, regardless of its distance from the camera. Further analysis is shown in Fig. 7, which features time-series plots of the mirror angles and zoom ratios for feed-forward control,

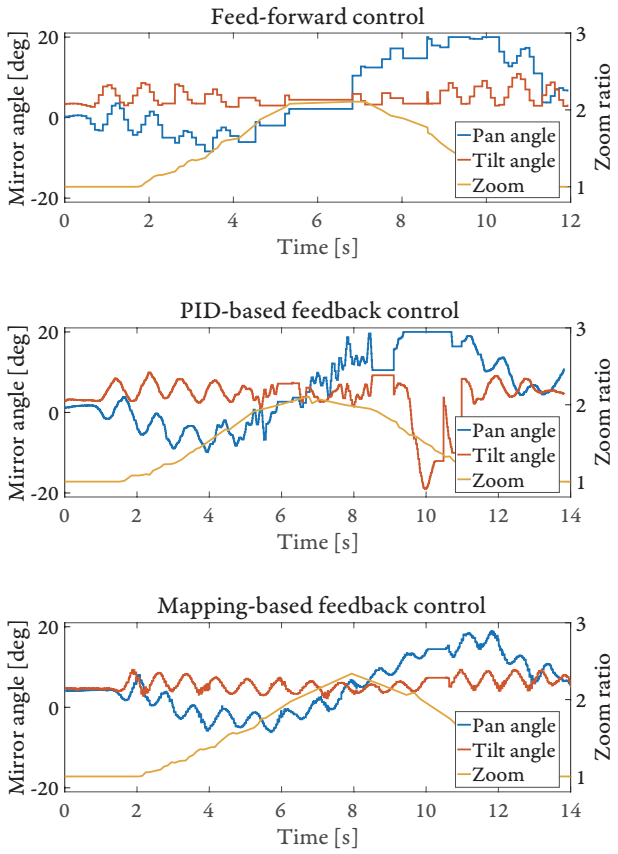

Fig. 7. Comparison of the pan, tilt angles, and zoom ratio for three different control strategies. The pan and tilt angles correspond to the left axis and the zoom ratio corresponds to the right axis.

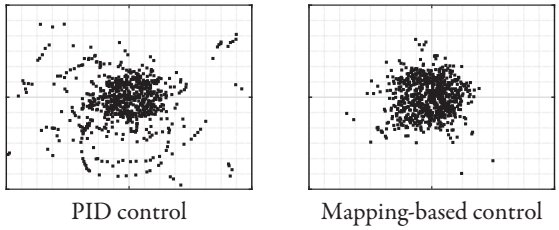

Fig. 8. Distributions of detected object's location on the image.

PID-based feedback control, and the proposed mapping-based feedback control during ball tracking. These results demonstrate smoother and more robust tracking with our proposed visual feedback control compared to both feed-forward and PID-based controls. We observed a small vibration around the 4-second mark due to the slowness of the zoom lens, which we aim to rectify by introducing a faster variable zoom lens. In Fig. 8, we compared the detected locations of the ball between our proposed mapping-based feedback control and conventional PID-based feedback. The former consistently tracked the ball nearer to the center of the image. We anticipate that with higher image acquisition rates and improved processing times, the object's center location distribution could be further minimized, enabling tracking of faster targets. This remains a focus for future work.

V. CURVED MIRROR TRACKING

A. Setting

We also explored the Falcon's utility in intelligent vehicles by testing curved mirror tracking on public roads. Curved mirrors, commonly used in Japan to enhance driver visibility, were our primary focus. We securely affixed the prototype atop a passenger vehicle roof, 1.85 m above the ground, using helical anti-vibration springs (Fig. 1). We preset the zoom ratio to 3.0 and focus position to 30.0 m, then drove around Tokyo to evaluate system performance.

B. Training

To train a deep neural network for curved mirror recognition, we compiled a dataset of 160 images captured from public roads, with each image manually annotated for the curved mirror. We supplemented this with an online public dataset containing 800 curved mirror images. We trained a deep neural network on YOLOv8-s [12] using transfer learning, striking a balance between inference time and accuracy. The network was compiled using TensorRT, consistent with our previous evaluation.

C. Results

Fig. 9 demonstrates the ability of Falcon to stably track a curved mirror, with the visual feedback from our proposed method centering the mirror in the image. A comparison of images taken by our system and an external camera is presented in Fig. 10. We confirmed clear visibility of the situation behind the corner in the curved mirror's reflection. Despite the external camera's high resolution (2880×1860 px), the imagery of Falcon surpassed it in terms of resolution, which enhanced subsequent processes such as pedestrian detection in blind spots, thereby promoting safer intelligent vehicles. Fig. 11 depicts the continuous control of galvanometer mirrors and the zoom lens to maintain focus on the curved mirror.

VI. CONCLUSION

In this study, we proposed an innovative active vision system composed of an electric zoom lens attached to a high-speed camera and fast-tilting mirror pair for gaze control. Furthermore, we introduced a novel, accurate, mapping-based calibration method for Falcon, integrated with external cameras. We also introduced a mapping-based feedback control method that enabled robust object tracking with minimal computational overhead. Validated in an indoor environment, our system demonstrated superior performance over conventional mirror-control methods. Moreover, we demonstrated the effectiveness of the system in tracking a curved mirror on public roads, with Falcon outperforming commercial high-resolution onboard cameras in image quality. Given its versatility, the proposed system showed potential

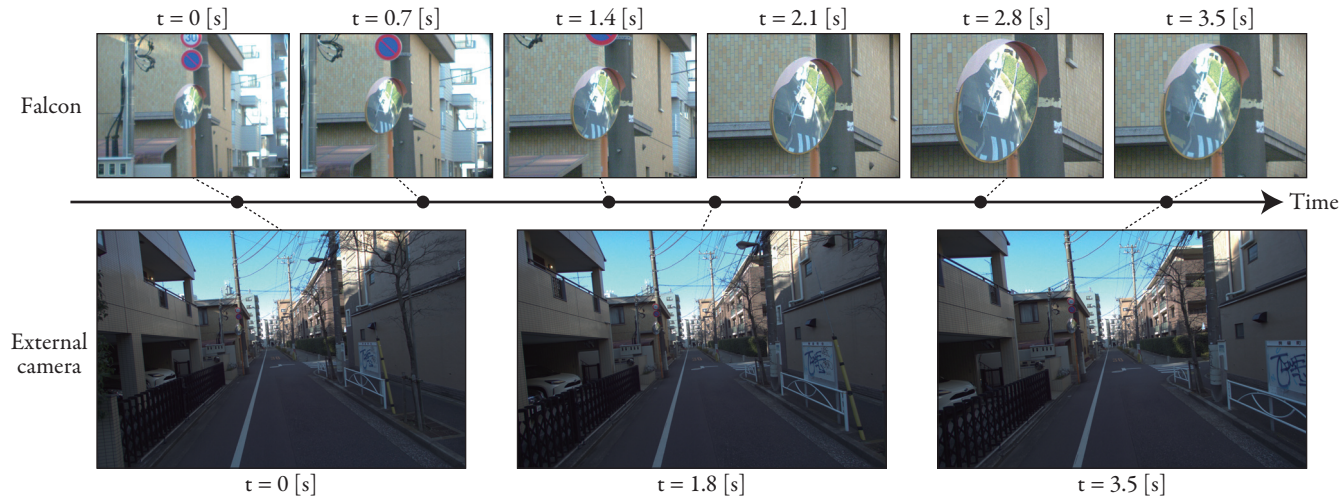

Fig. 9. Imagery by the Falcon and the external camera.

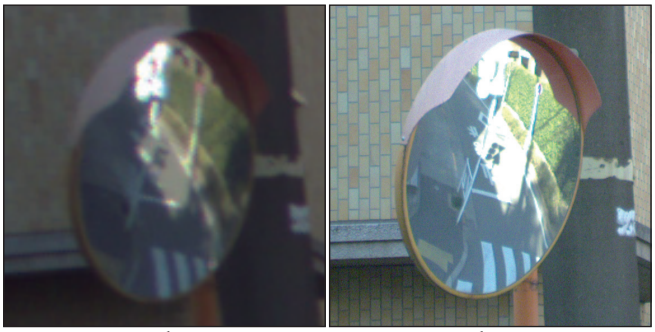

Fig. 10. Comparison of the images of the curved mirror. Both images were captured at $t = 3.5$[s].

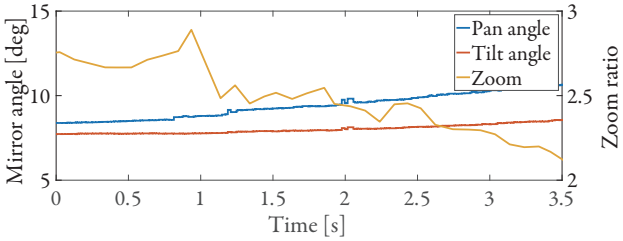

Fig. 11. Pan, tilt angles and zoom ratio at curved mirror tracking.

for applications beyond automobiles, including mobile robots and surveillance cameras. Currently, the system's size and weight are substantial, so we plan to incorporate smaller components for practical use. Future work will also explore developing practical ITS applications using Falcon, such as traffic signal recognition and observation of dangerous spots.

References

[1] A. W. Bergman, D. B. Lindell, and G. Wetzstein, "Deep adaptive lidar: End-to-end optimization of sampling and depth completion at low sampling rates," in *International Conference on Computational Photography*, 2020, pp. 1–11.

[2] F. Pittaluga, Z. Tasneem, J. Folden, B. Tilmon, A. Chakrabarti, and S. J. Koppal, "Towards a mems-based adaptive lidar," in *International Conference on 3D Vision*, 2020, pp. 1216–1226.

[3] T. Sueishi, H. Oku, and M. Ishikawa, "Lumipen 2: Dynamic Projection Mapping with Mirror-Based Robust High-Speed Tracking against Illumination Changes," *Presence: Teleoperators and Virtual Environments*, vol. 25, no. 4, pp. 299–321, 2016.

[4] T. Sueishi, M. Ishii, and M. Ishikawa, "Tracking background-oriented schlieren for observing shock oscillations of transonic flying objects," *Applied Optics*, vol. 56, no. 13, pp. 3789–3798, 2017.

[5] B. Tilmon and S. J. Koppal, "Saccadecam: Adaptive visual attention for monocular depth sensing," in *International Conference on Computer Vision*, 2021, pp. 5989–5998.

[6] Q. Li, M. Chen, Q. Gu, and I. Ishii, "A flexible calibration algorithm for high-speed bionic vision system based on galvanometer," in *International Conference on Intelligent Robots and Systems*, 2022, pp. 4222–4227.

[7] T. Zhang, S. Hu, K. Shimasaki, I. Ishii, and A. Namiki, "Dual-camera high magnification surveillance system with non-delay gaze control and always-in-focus function in indoor scenes," in *International Conference on Intelligent Robots and Systems*, 2022, pp. 6637–6642.

[8] K. Okumura, H. Oku, and M. Ishikawa, "High-speed gaze controller for millisecond-order pan/tilt camera," in *International Conference on Robotics and Automation*, 2011, pp. 6186–6191.

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

[10] K. Iida and H. Oku, "Saccade mirror 3: High-speed gaze controller with ultra wide gaze control range using triple rotational mirrors," in *International Conference on Robotics and Automation*, 2016, pp. 624–629.

[11] B. Tilmon, E. Jain, S. Ferrari, and S. Koppal, "Foveacam: A mems mirror-enabled foveating camera," in *International Conference on Computational Photography*, 2020, pp. 1–11.

[12] Ultralytics, "YOLOv8." [Online]. Available: https://github.com/ultralytics/ultralytics

[13] NVIDIA, "TensorRT." [Online]. Available: https://developer.nvidia.com/tensorrt