2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Dynamic-Range Focal Sweep: Seamless Continuous Autofocus Based on High-Speed Vision for Magnified Object Tracking

Tianyi ZHANG[1], Kohei SHIMASAKI[2], Idaku ISHII[2] and Akio NAMIKI[1]

***Abstract*— This paper presents an innovative continuous autofocus (C-AF) approach based on high-speed vision. It consistently provides focused images with stable and sufficiently high frame rates, aiming to improve the ability to track small, fast-moving objects in a highly magnified scene. To achieve this, we propose the concept of a dynamic-range focal sweep enabled by a high-speed camera and a focus-tunable liquid lens with high adjustment capability. The focal sweep consistently covers a small range around the object's focus position, guided by previous depth results obtained through the depth-from-focus (DFF) technique. We conducted verification experiments to thoroughly analyze the capability of the proposed C-AF approach. By integrating a 2-axis Galvano mirror, we built a high-speed C-AF active vision system with rapid focus and pan-tilt adjustments. The comprehensive experiment results highlight the advanced capabilities of our seamless C-AF in tracking magnified objects.**

I. INTRODUCTION

Active vision systems (AVS) are widely recognized for their ability to capture high-resolution images of objects of interest while dynamically adjusting the viewing direction to track their motion, and they are used in diverse fields such as automated manufacturing [1], intelligent surveillance [2], and smart agriculture [3]. However, tracking high-magnified, fast-moving small objects poses significant challenges due to the constraints of narrow field-of-view (FoV) and shallow depth-of-field (DoF), making rapid pan and tilt adjustments and maintaining focus crucial to prevent tracking failures.

While some new pan-tilt mechanisms have improved the viewing adjustment ability [4], [5], [6], [7], autofocus (AF) technology faces limitations like low precision and high delays. These cause continuous autofocus (C-AF) difficult, as tracking tasks require a steady stream of well-focused images with sufficiently high frame rates.

In general, AF approaches can be classified into active AF and passive AF [8]. Active AF employs depth sensors to measure object's distance so that focus adjustment can be completed with only one step [9], [10], [11]. Whereas, they encounter challenges with small objects due to potential inaccuracies in measuring distances, stemming from the limitations of sensor resolution and signal attenuation.

*This work was supported by JST SPRING, Grant Number JP-MJSP2019

[1]Tianyi ZHANG, Akio NAMIKI are with the Namiki Laboratory, Dept. of Mechanical Engineering, Graduate School of Engineering, Chiba University, Japan zty2023@gmail.com, namiki@faculty.chiba-u.jp

[2]Kohei SHIMASAKI, Idaku ISHII are with the Smart Robotics Laboratory, Graduate School of Advanced Science and Engineering, Hiroshima University, Japan simasaki@hiroshima-u.ac.jp, iishii@robotics.hiroshima-u.jp

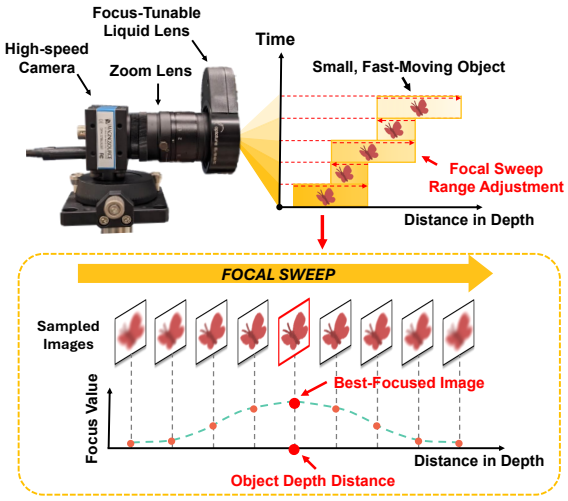


Fig. 1. Dynamic-Range Focal Sweep-Based C-AF Approach: Images at different focuses are captured during the focal sweep. The focus measure algorithm extracts the best-focused image as the output. Meanwhile, the DFF technique measures the object's depth, which is then used to adapt the range of subsequent focal sweeps. This ensures that the object remains within the focal sweep range consistently.

On the other hand, passive AF can branch into phase-based AF and contrast-based AF. The former uses phase sensors for focus detection [12], [13], [14], while the latter adjusts focus progressively using search methods [15]. Although some predictive algorithms, such as depth from defocus [16], [17], and machine learning [18], [19], [20], effectively reduced iterative steps compared to traditional methods like rule-based search [21] and fast climbing search [22], using passive AF for C-AF remains challenging for shallow DoF conditions due to unacceptable delays, as refocusing process has to start after detecting incorrect focus.

Recently, faster opto-mechanisms, including piezoelectric actuators [23], voice coil motors [24], and focus-tunable liquid lenses [25], have been developed, accelerating focus adjustment to mere tens of milliseconds per step. This facilitates the AF based on global focal sweeps, which obtains the focused images with a stable frame rate by extracting the best-focused one from couples of images captured with varying focuses [26], [27], [28], [29], [30]. However, global focal sweep produces vast redundancy of frames, leading to low output frame rates.

In this study, we present the concept of dynamic-range focus sweep for C-AF illustrated in Fig. 1. This high-speed AF approach can obtain the focused images of high-magnified objects consistently with sufficiently high frame

rates for object tracking, by adjusting the focal sweep range seamlessly based on the object depth measured through Depth-from-Focus (DFF) technique [31], [32]. In our system, we achieve it by leveraging a 500 fps high-speed camera and a focus-tunable liquid lens to perform 50 Hz focal sweeps. As a result, 10 sampled images are captured in each sweep, and the image with the highest focus value is extracted as the output. Unlike global focal sweep, our approach effectively reduces the redundancy of sampled images by narrowing the focal sweep range to a short depth range around the object's true focus, ensuring the well-focused output with stable 50 fps. We completely analyzed the C-AF performance of our method. Furthermore, we designed a high-speed AVS by combining the proposed C-AF camera with a Galvano-mirror-based pan-tilt mechanism. The experiment results showcase its advanced ability for magnified object tracking.

II. Continuous Autofocus Based on Dynamic-Range Focal Sweep

A. Focal Sweep with High-Speed Vision

Given the high speed of the focal sweep driven by the liquid lens and high-speed camera, changes in the object during the sweep can be negligible. Thus, we depict the schematic of the focal sweep in Fig 2.

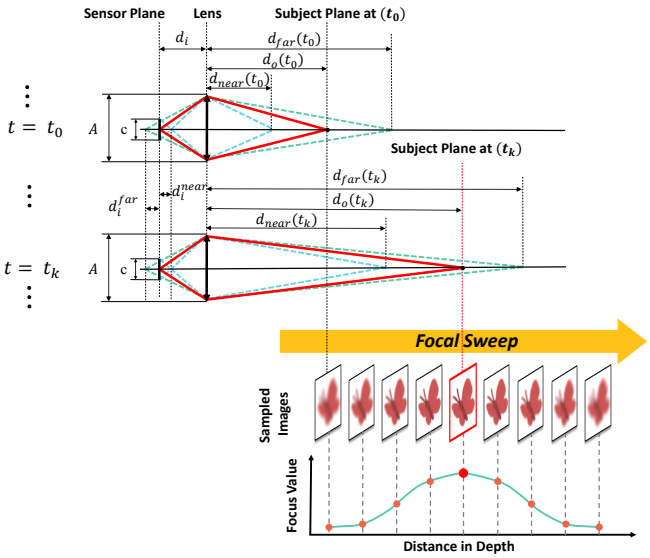


Fig. 2. Schematic of Focal Sweep: The upper part illustrates the optical geometry based on the Thin Lens Law, while the lower part describes the sampled images and their focus values obtained during the focal sweep.

First, the relationship between the focal length f, the camera-lens distance d_i, and the object-lens distance d_o adheres to Gaussian Lens Equation 1, expressed as:

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}. \tag{1}$$

Here, the focal length f is determined by the equation $\frac{1}{f} = \frac{1}{f_z} + \frac{1}{f_l}$, where f_z and f_l represent the focal lengths of the zoom lens and the liquid lens, respectively.

According to the intercept theorem, the nearest and farthest distances between the lens and the image plane that capture a focused image, denoted as d_i^{near} and d_i^{far}, can be expressed in Equation 2,

$$d_i^{near} = \frac{d_i}{1 - \frac{c}{A}}, \quad d_i^{far} = \frac{d_i}{1 + \frac{c}{A}}, \tag{2}$$

where A refers to the aperture diameter, and c refers to the acceptable circle of confusion (CoC).

Using Equation 1 and 2, when the image is located at d_o, the nearest and farthest distances d_{near} and d_{far} of the DoF can be calculated through Equation 3:

$$d_{near} = \frac{Afd_o}{Af + c(d_o - f)}, \quad d_{far} = \frac{Afd_o}{Af - c(d_o - f)}. \tag{3}$$

To ensure that at least one focused image is captured during the focal sweep, we impose the condition that the nearest distance of the next sampling instance, $d_{near}(t_{k+1})$, should be less than or equal to the farthest distance of the current sampling instance, $d_{far}(t_k)$. In practice, the range of the focal sweep is set manually, considering that the acceptable CoC is task-dependent and can vary widely across different computer vision applications.

Additionally, in our work, the focus value is defined as the average of the pixel values of the edge maps that obtained through the GPU-accelerated Canny sharpness operator [33]. The highest focus value in its focal sweep process is extracted as the output image.

B. Dynamic-Range Focal Sweep

We implement the depth measurement method based on the DFF technique [31]. As described in Fig. 3, the focal sweep is operated by controlling the liquid lens' diopter, i.e., focal power, denoted as δ ($\delta = \frac{1}{f_l}$). This control is executed in a sinusoidal mode by sending commands for the maximal and minimal focal powers, δ^{max} and δ^{min}, respectively.

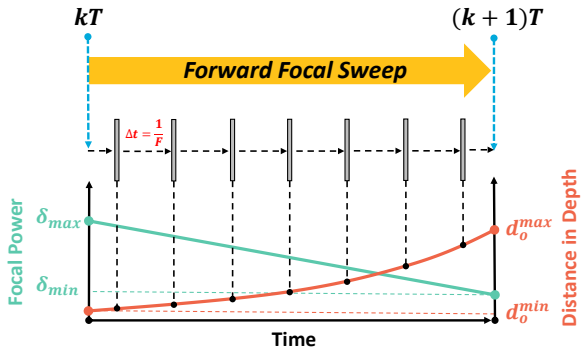


Fig. 3. Schematic of Depth Measurement with Focal Sweep.

Owing to the high sweep frequency, the variation of focal power is simplified as uniform. Thus, during any forward focal sweep, the focal power at time t can be calculated using Equation 4,

$$\delta(t) = \delta_{max} - \frac{\delta_{max} - \delta_{min}}{T} t \quad (0 \le t \le T), \tag{4}$$

where T is the period of the focal sweep.

In a forward focal sweep with given maximal and minimal positions, denoted as d_o^{max} and d_o^{min}, the relationship between the focal position and the sampling image capture time is expressed by Equation 5:

$$d_o(t) = \frac{d_o^{max} d_o^{min} T}{(d_o^{min} - d_o^{max})t + d_o^{max} T} \quad (0 \le t \le T). \qquad (5)$$

Fig. 4 illustrates the process of dynamic-range focal sweep. There are three sequential focus sweep processes in chronological order: from kT to $(k+1)T$, from $(k+1)T$ to $(k+2)T$, and from $(k+2)T$ to $(k+3)T$. The first and third processes are forward, while the second process is backward.

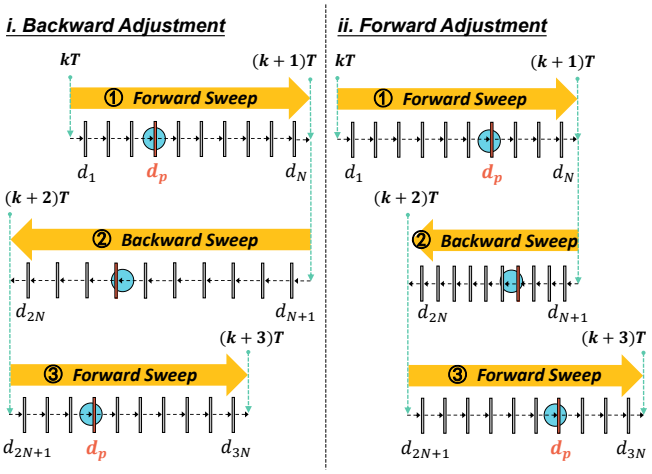


Fig. 4. Diagram of Dynamic-Range Focus Sweep: Sub-figures *i.* and *ii.* illustrate the adjustment of the focal sweep's range as the object moves backward and forward, respectively. In the first sweep, the object's distance is measured. It is used to adjust the range in its subsequent focal sweeps.

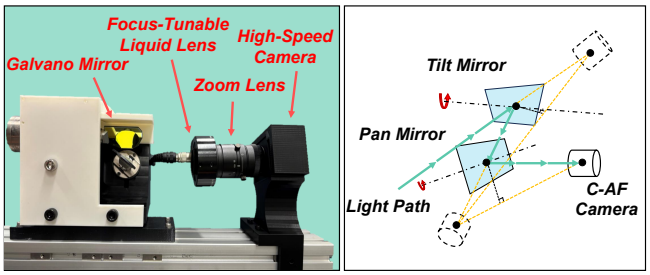
In each process, the best-focused image is extracted. Since the backward focal sweep is unstable due to range adjustments, focus positions d_p are measured only during forward sweeps using Equation 5.

C. System Synchronization

Since the liquid lens operates in an open-loop, an initialization process is required to determine the start time and direction of each focal sweep. Here, we applied the following method. First, we roughly estimate the object's depth using a single global focal sweep with the liquid lens in its current mode. Second, we start the sine wave mode around the object's depth. Due to the high sweep frequency, the object movement is negligible. Thus, the start time of the second focal sweep, denoted as t_s, can be obtained with $t_s = \frac{t_1+t_2}{2}$, where t_1 and t_2 are the capture time of the first and the second best-focused images. Since the focal sweep period T is constant, the start time of the k-th sweep is $t_s + (k-1)T$.

To distinguish the sweep directions, we move the focal sweep farther, and calculate the time variation of the capture time of the best-focused image in the forth focal sweep, denoted as Δt_4. It can be calculated by $\Delta t_4 = t_4 - (t_2 + 2T)$, where t_4 is the capture time of the forth best-focused image. If $\Delta t_4 < 0$, the forth focal sweep is forward; otherwise, the forth focal sweep is backward.

III. System Setup



(a) (b)

Fig. 5. AVS Using the Proposed C-AF Camera and a 2-axis Galvano Mirror: Sub-figure (a) shows the hardware configuration of the AVS combining the Galvano mirror system for pan-tilt adjustment and the proposed C-AF camera for obtaining the focused images of the tracked object. Sub-figure (b) illustrates the optical geometry of the Galvano mirror, with details available in previous works [6], [7].

As shown in Fig. 5, the AVS system consists of a 2-axis Galvano mirror (6210H, Cambridge Technology) and a C-AF camera. The Galvano mirror adjusts the horizontal and vertical directions of the view within a 40° FoV within 2 ms. The C-AF camera is composed of a focus-tunable liquid lens (EL-16-40-TC-VIS-20D-1-C, Optotune) controlled by the Optotune Lens Driver 4i, a 100 mm focal length zoom lens, and a high-speed camera (DFK27BUX287, ImagingSource). The camera captures 720x540 resolution images at 500 fps. The liquid lens operates in the sine wave mode at the frequency of 25 Hz, generating 50 Hz focal sweeps, resulting in 10 sampled images per sweep.

IV. Verification Experiments

In this section, we conducted two experiments to verify the performance of the proposed C-AF approach. Sub-section A introduces the experimental environment setup. Sub-section B and Sub-section C evaluate the depth measurement accuracy and AF performance during the C-AF task with multiple moving objects under different lighting conditions, respectively.

A. Experimental Environment Setup

As shown in Fig. 6, a linear actuator was used to make the objects move between 0.5 m and 2.0 m. The objects were shown in sub-figures (a)-(f). Their information is listed in Table I.

TABLE I
Detailed Information of Objects

Object No.	Object Item	Width (cm)	Length (cm)
(a)	Alphabet Letters	0.9	0.9
(b)	Kanji Characters	1.0	1.0
(c)	QR Code	2.0	2.0
(d)	ArUco Marker	2.0	2.0
(e)	Butterfly Model	2.0	2.0
(f)	Screw	1.5	1.0

In order to demonstrate the robustness of our method, we conducted the experiment to evaluate the depth measurement

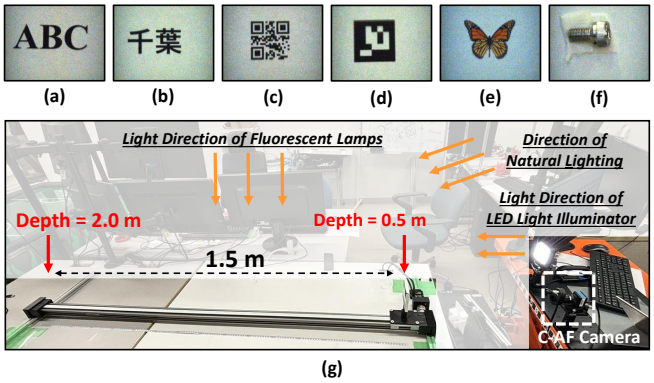


Fig. 6. Experiment Setup.

accuracy and focus accuracy under three distinct lighting scenarios, in the order from bright to dark, are:

- Lighting Condition (i): Fluorescent Lamps + LED Light Illuminator;
- Lighting Condition (ii): Indoor natural lighting with fluorescent lamps;
- Lighting Condition (iii): Indoor natural lighting.

B. Depth Measurement Accuracy Analysis

This experiment evaluated the depth measurement accuracy during the object's movements. Here, the objects were moved between 0.5 m and 2.0 m at 3.0 m/s. Fig. 7, Fig. 8, and Fig. 9 show the depth measurement results under different lighting conditions, respectively. In these figures, the objects' backward movements are during the time intervals from the 0.5 s to the 1.0 s; similarly, their forward movements are from the 1.5 s to the 2.0 s.

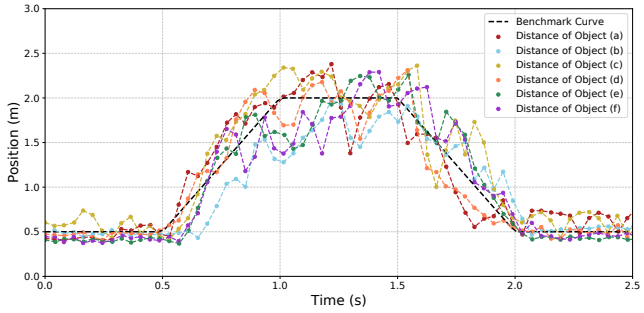


Fig. 7. Depth Measurement Results under Lighting Condition (i).

We calculated the measurement error to quantify the accuracy of depth measurement. Specifically, given that the objects were moved at a uniform speed of 3 m/s and our C-AF approach measures depth at a constant rate of 25 Hz, we assessed the measurement error for distances of 0.75 m, 1.25 m, and 1.75 m using the results from the corresponding frame numbers. The detailed results are listed in Table II.

It is evident that the depth measurement is both effective and continuous during object movement. Depth measurement accuracy improves when objects are closer, as the focal sweep range can be smaller due to the narrower DoF for each sampled image. Conversely, accuracy decreases at greater

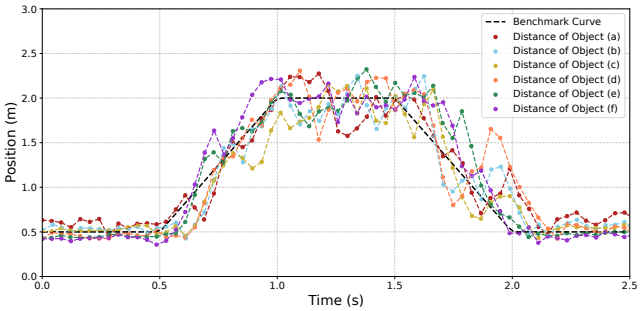


Fig. 8. Depth Measurement Results under Lighting Condition (ii).

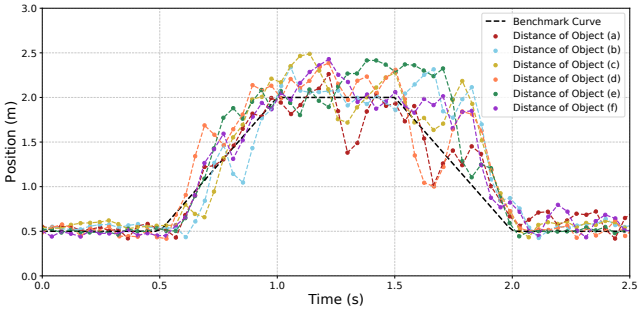


Fig. 9. Depth Measurement Results under Lighting Condition (iii).

TABLE II
DEPTH MEASUREMENT ERROR AT DIFFERENT DISTANCES (UNIT: M)

	Lighting Condition (i)			Lighting Condition (ii)			Lighting Condition (iii)		
	0.75 m	**1.25 m**	**1.75 m**	**0.75 m**	**1.25 m**	**1.75 m**	**0.75 m**	**1.25 m**	**1.75 m**
(a)	0.261	0.505	0.174	0.170	0.040	0.025	0.082	0.032	0.124
(b)	0.207	0.159	0.342	0.402	0.182	0.281	0.206	0.473	0.261
(c)	0.074	0.197	0.364	0.221	0.183	0.323	0.112	0.502	0.138
(d)	0.123	0.215	0.421	0.557	0.213	0.175	0.139	0.403	0.520
(e)	0.145	0.260	0.157	0.091	0.313	0.142	0.121	0.279	0.443
(f)	0.183	0.344	0.202	0.022	0.064	0.321	0.056	0.465	0.211
Avg.	0.165	0.280	0.276	0.244	0.166	0.211	0.119	0.359	0.283

distances because a longer focal sweep range is required to accommodate the wider DoF. Despite this, the C-AF task was still successfully completed. Additionally, although darker conditions challenge depth measurement precision and make it difficult to extract the best-focused image, the results demonstrate that the proposed C-AF approach can accurately measure depth under varying lighting conditions.

C. Autofocus Performance Analysis

In this experiment, we conducted the subjective and objective evaluations to demonstrate the AF performance of the proposed C-AF approach during the object's movements at 3 m/s.

1) Subjective Evaluation: We showed images obtained under the lighting condition (i) in Fig. 10. Here, the benchmark images and the outputs are shown in rows (1) and (2), respectively. The method to obtain these outputs was the same as the method employed in the previous experiment. Moreover, in order to showcase the focus details, the zoomed details of the objects are shown in the row (3).

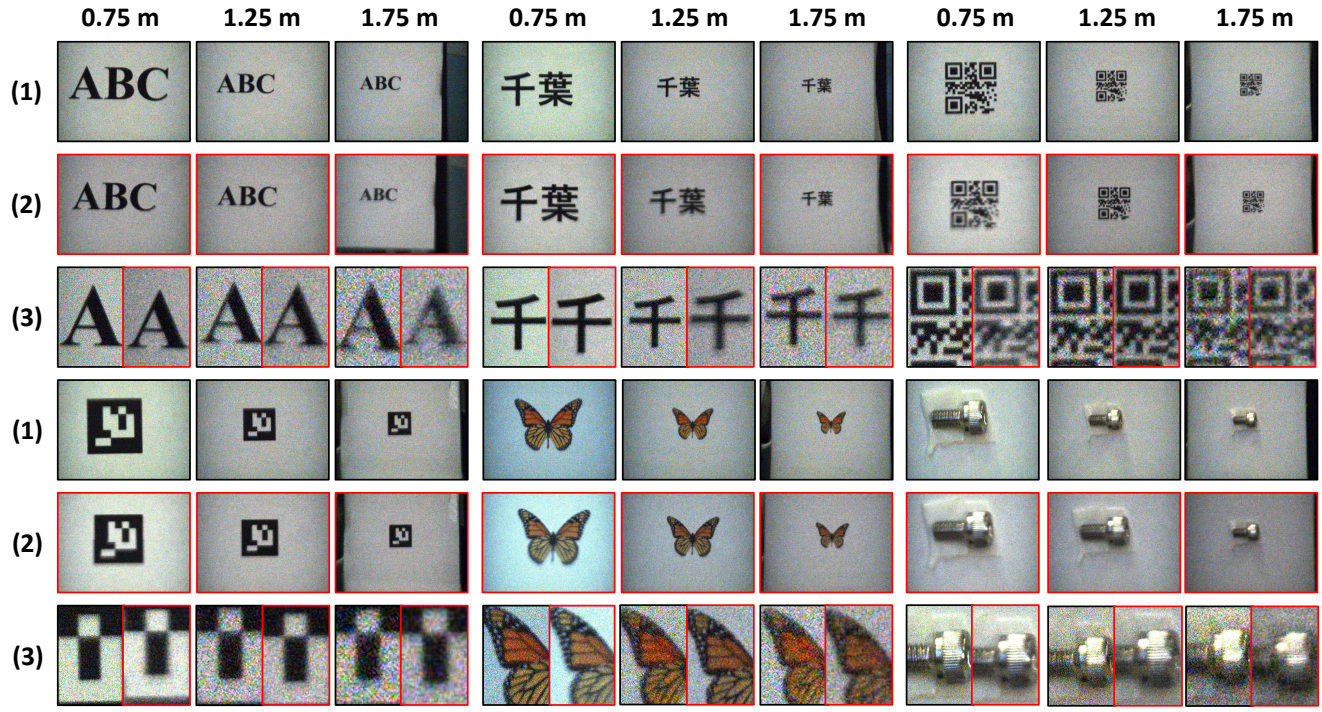


Fig. 10. Comparison between C-AF Outputs and Images with Optimal Focus: Row (1) presents the benchmark images, captured when the objects were stationary at their corresponding locations, ensuring optimal focus. Row (2) displays the outputs captured by our system during the objects' movements. Finally, Row (3) includes zoomed-in sections of the objects, providing a visual comparison of the details to highlight the focus differences between the C-AF outputs and the benchmark images.

It is easy to see that the outputs exhibit similar focus states to the optimal ones, as major shapes and edges are distinguishable. However, a slight gap remains in capturing finer object details compared to the optimal images. This discrepancy can be attributed to two main factors:

- Firstly, although the dynamic-range focal sweep significantly reduces the depth gap between sampled images, a residual depth gap persists between the optimal and sampled depths.
- Secondly, due to the significant noise in images captured by the high-speed camera, a Gaussian filter was applied for denoising. This pre-processing step, while effective for noise reduction, inevitably results in a loss of some object details.

2) Objective Evaluation: In order to quantify the focus performance objectively, we evaluate the focus difference and calculate the success rate under different robot vision tasks.

First, we defined the Focus Value Relative Loss (*FVRL*), which is calculated as $FVRL = \frac{v-\hat{v}_k}{v} \cdot 100\%$, where v represents the focus value of the optimal image, and $\hat{v}_k$ denotes the focus value of the output image. The focus value is determined by summing the edge pixels of the edge map generated using the Canny sharpness operator, with minimum and maximum thresholds set to 60 and 180. To resolve the strong noise in the images captured by the high-speed camera, a Gaussian filter with a kernel size of (5,5) and a sigma value of 3 was applied. This algorithm is the same as the focus measure algorithm that we used to extract the outputs in our system. The results are shown in Table III.

TABLE III
FOCUS VALUE RELATIVE LOSS AT DIFFERENT DEPTH DISTANCES

	Lighting Condition (i)			Lighting Condition (ii)			Lighting Condition (iii)		
	0.75 m	**1.25 m**	**1.75 m**	**0.75 m**	**1.25 m**	**1.75 m**	**0.75 m**	**1.25 m**	**1.75 m**
(a)	23.63%	0.00%	55.11%	3.39%	78.32%	30.67%	20.75%	0.00%	20.96%
(b)	0.47%	68.89%	35.65%	10.01%	72.97%	24.93%	0.00%	48.49%	52.70%
(c)	69.62%	9.63%	23.35%	60.07%	1.79%	84.11%	21.58%	2.11%	0.00%
(d)	50.45%	15.26%	18.38%	62.57%	61.21%	59.00%	14.08%	14.35%	13.80%
(e)	62.37%	16.42%	67.71%	32.92%	61.44%	51.47%	29.59%	20.04%	34.60%
(f)	70.36%	31.01%	53.44%	47.40%	41.86%	26.35%	27.01%	64.20%	60.60%
Avg.	46.15%	23.54%	42.27%	35.50%	52.93%	46.10%	18.84%	24.87%	30.44%

As shown in Table III, while there is consistently a discrepancy between the output and the benchmark images with optimal focus, the major edges are preserved in the outputs across different distances. The detailed analysis is as follows:

- In bright lighting conditions, fine edge details are distinguished more easily, so any blur leads to greater loss value. Conversely, in darker conditions, where fewer details are visible, the value is lower since fine edges are less detectable in both the benchmark and output images. Although the low-light scenes may lead to challenges, in every condition, the proposed C-AF approach showcased the C-AF capability because the major edges remained.
- For objects with rich edge details, the loss value is higher because more edges contribute to the loss when blur occurs. Simpler objects, with fewer edges, exhibit a

lower value, as fewer details are impacted by blurriness, whereas this also leads to the imprecise best-focused image extraction using our C-AF approach.

- At closer distances, the abundance of edge pixels contributes to greater loss when blur is present. However, at greater distances, the value also increases due to the increased blur caused by the wider focal sweep range required to capture images, especially in bad lighting conditions.

In addition, to demonstrate the effectiveness of the proposed C-AF approach in consistently tracking magnified objects for various robotics tasks, we implemented several common computer vision algorithms. Specifically, as shown in Table IV, we applied optical character recognition algorithms to Objects (a) and (b), marker recognition algorithms to Objects (c) and (d), and object detection algorithms to Objects (e) and (f).

TABLE IV
DETAILED INFORMATION AND SUCCESS RATES WITH VARIOUS COMPUTER VISION ALGORITHMS

Object No.	Task Item	Implemented Algorithm	Success Rate with The Proposed AF
(a)	Optical Character Recognition	EasyOCR [34]	94.93%
(b)	Optical Character Recognition	EasyOCR [34]	93.83%
(c)	QR Code Recognition	QReader [35]	75.88%
(d)	ArUco Marker Recognition	ArUco Module in OpenCV [36]	97.63%
(e)	Object Detection	YOLOv8 [37]	99.25%
(f)	Object Detection	YOLOv8 [37]	78.93%

The results demonstrate that our C-AF system effectively detects most objects in the outputs. The highest success rate of 99.25% was achieved for object detection in Object (c), while the lowest, 75.88%, was observed for QR code recognition. This data suggests that the acceptable defocus varies across tasks, indicating room for further improvement in the camera system. However, the overall robustness and effectiveness of the proposed C-AF approach are evident. The majority of outputs were sufficiently well-focused to allow the algorithms to accurately detect and extract detailed visual information. Even in the highly challenging task of QR code recognition, the successful detection underscores the C-AF system's ability to continuously obtain precisely focused images during object movement. This is particularly noteworthy given the challenges posed by the very narrow DoF in high-magnified view and the high noise in images captured by high-speed cameras.

V. COMPREHENSIVE EXPERIMENT

In this section, we conducted a comprehensive experiment to demonstrate the magnified object tracking performance using the high-speed AVS combining the proposed C-AF approach with high-speed pan-tilt mechanism based on a 2-axis Galvano mirror.

A. Experimental Environment Setup

The experimental setup is illustrated in Fig. 11. In this experiment, the butterfly model (Object (d)) and the QR code (Object (e)) move along the track paths marked in purple and blue, respectively. Both motions started at position *A*, which is located directly in front of the camera's position *O*.

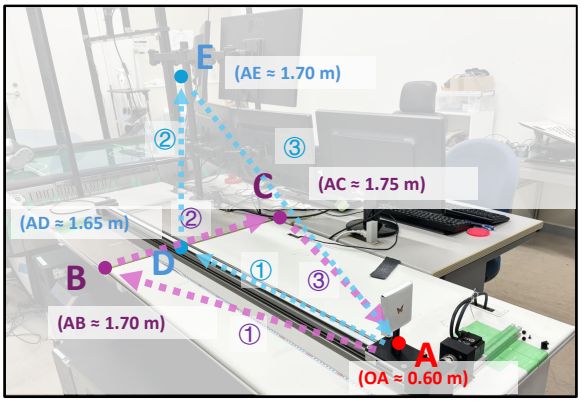


Fig. 11. Experimental Environment Setup for Magnified Object Tracking.

Once the motion begins, the proposed C-AF approach captures well-focused outputs at 50 fps seamlessly. In this experiment, the CSR-DCF tracking algorithm [38] is implemented to calculate the displacement between the object's center and the image center. This data is used to adjust the pan-tilt angles, ensuring the object remains consistently within the magnified view.

B. Experimental Result

We selected outputs at 0.5-second intervals for display in this paper. As shown in Fig. 12, these images remain consistently clear, enabling the object tracking algorithm to perform accurate image tracking.

Fig. 12. Magnified Output Images Captured During Objects' Movements.

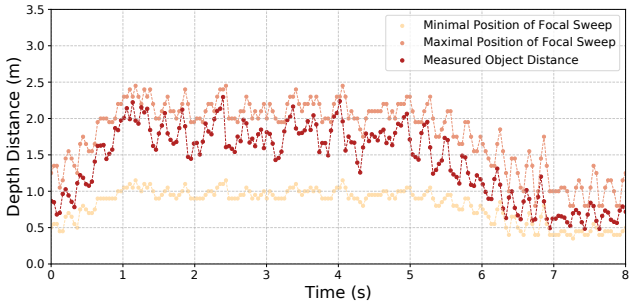


Fig. 13. Variation of Focal Sweep Range For Tracking the Butterfly Model.

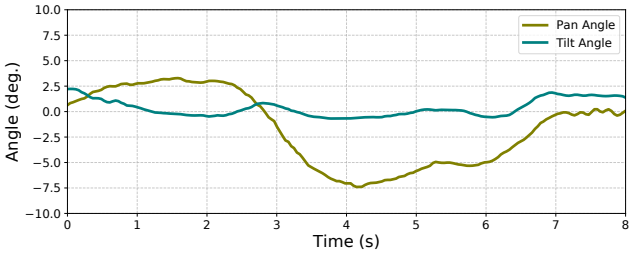


Fig. 14. Galvano Mirror's Pan-Tilt Angles During the Butterfly Model's Movement.

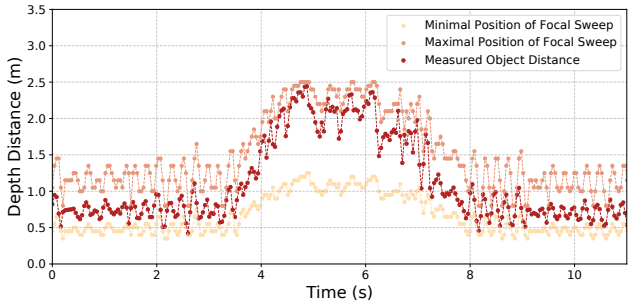


Fig. 15. Variation of Focal Sweep Range For Tracking the QR Code.

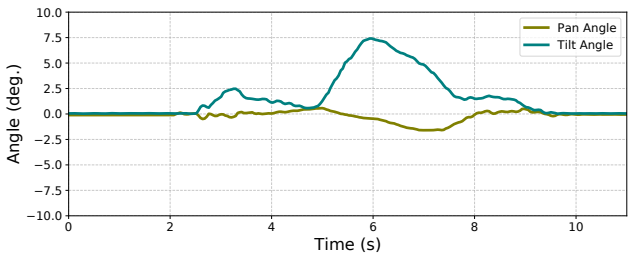


Fig. 16. Galvano Mirror's Pan-Tilt Angles During the QR Code's Movement.

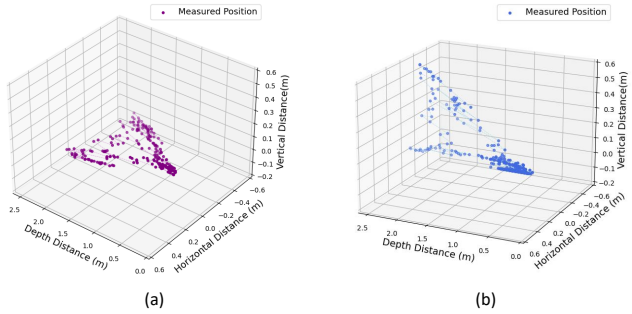


Fig. 17. 3D Positions of the Objects During the Movements.

Fig. 13 and 15 include three curves representing the maximum and minimum positions of the focal sweep range, as well as the measured position calculated using the proposed C-AF approach. Although there is still room for improvement in depth measurement accuracy, our method effectively adjusts the focal sweep based on the object's approximate depth, enabling seamless C-AF operation.

Consequently, we obtain well-focused outputs with a stable and sufficiently high frame rate, which is crucial for object tracking in magnified scenes. As depicted in Fig. 14 and 16, our output images are clear enough to accurately determine the object's displacement, and the high frame rate ensures that the Galvano mirror can smoothly adjust its pan-tilt directions.

As a result, as illustrated in Fig. 17, by using the measured depths and pan-tilt angles, 3D tracking of magnified objects is realized, demonstrating the approach's strong performance and potential for robotic tasks in real-world applications.

VI. Discussion

Our study introduces a novel C-AF approach aimed at addressing the unpredictable intervals of blur encountered in existing C-AF methods. The core innovation in this paper, dynamic-range focal sweep, effectively enables a sufficient and stable frame rate of well-focused output. While it offers significant advantages, it is important to acknowledge its limitations. Firstly, the computation of high-frame-rate images in real time demands substantial computational resources. Secondly, the open-loop liquid lens driver requires an initialization process to achieve system synchronization. Therefore, future research should focus on implementing closed-loop control and prediction algorithms to reduce the required frame rate of sample images.

VII. Conclusion

This study presents an innovative C-AF approach based on high-speed vision. The method employs a high-speed camera and a liquid lens that dynamically adjusts the focal sweep around the object's true focus. The sweep range is adaptable, informed by depth measurements using the DFF technique, eliminating the need for additional depth sensors. This approach achieves seamless C-AF, even with narrow DoF and rapid object movements. In our experiments, we achieved stable, well-focused images at 50 fps by implementing a 50 Hz dynamic focal sweep using a 500 fps high-speed camera. To validate the proposed C-AF approach, we conducted experiments involving subjective evaluation and quantitative analysis with various objects under different lighting conditions. Additionally, we developed a C-AF Active Vision System (AVS) to demonstrate the approach's potential in magnified object tracking. The results indicate that the system reliably measures distances and adjusts the focal sweep range in response to changes in object depth, ensuring that fast-moving objects remain in focus.

REFERENCES

[1] Q. Tarar and F. Mohr, "Optical Zoom System for Laser Welding Robot," 2007 17th International Conference Radioelektronika, Brno, Czech Republic, 2007, pp. 1-5, doi: 10.1109/RADIOELEK.2007.371703.

[2] L. Alam, M. M. Hoque, M. A. Akber Dewan, N. Siddique, I. Rano and I. H. Sarker, "Active Vision-Based Attention Monitoring System for Non-Distracted Driving," in IEEE Access, vol. 9, 2021, pp. 28540-28557, 2021, doi: 10.1109/ACCESS.2021.3058205.

[3] Q. Feng, W. Cheng, W. Zhang and B. Wang, "Visual Tracking Method of Tomato Plant Main-Stems for Robotic Harvesting, "2021 IEEE 11th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Jiaxing, China, 2021, pp. 886-890, doi: 10.1109/CYBER53097.2021.9588275.

[4] B. Tilmon, E. Jain, S. Ferrari and S. Koppal,, "FoveaCam: A MEMS Mirror-Enabled Foveating Camera," 2020 IEEE International Conference on Computational Photography (ICCP), St. Louis, MO, USA, 2020, pp. 1-11, doi: 10.1109/ICCP48838.2020.9105183.

[5] M. Jiang, K. Shimasaki, S. Hu, T. Senoo and I. Ishii, "A 500-Fps Pan-Tilt Tracking System With Deep-Learning-Based Object Detection," in IEEE Robotics and Automation Letters, vol. 6, no. 2, 2021, pp. 691-698, April 2021, doi: 10.1109/LRA.2020.3048653.

[6] K. Okumura, H. Oku and M. Ishikawa, "High-Speed Gaze Controller for Millisecond-Order Pan/Tilt Camera," 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 2011, pp. 6186-6191, doi: 10.1109/ICRA.2011.5980080.

[7] S. Hu, H. Dong, K. Shimasaki, M. Jiang, T. Senoo and I. Ishii, "Omnidirectional Panoramic Video System With Frame-by-Frame Ultrafast Viewpoint Control," in IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 4086-4093, April 2022, doi: 10.1109/LRA.2022.3150484.

[8] Y. Zhang, L. Liu, W. Gong, H. Yu, W. Wang, C. Zhao, P. Wang, and T. Ueda, "Autofocus System and Evaluation Methodologies: A Literature Review," Sensors and Materials, vol. 30, No.5, 2018, pp. 1165-1174.

[9] R. Abele, D. Fronte, P. Liardet, J. Boi, J. Damoiseaux and D. Merad, "Autofocus in infrared microscopy," 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), Turin, Italy, 2018, pp. 631-637, doi: 10.1109/ETFA.2018.8502648.

[10] H. Liang, K. Lu, X. Liu, and J. Xue, "The Auto-focus Method for Scanning Acoustic Microscopy by Sparse Representation," Sensing and Imaging, Vol. 20, no. 33, 2019, doi: 10.1007/s11220-019-0255-x.

[11] X. Zhang, F. Fan, M. Gheisari and G. Srivastava, "A Novel Auto-Focus Method for Image Processing Using Laser Triangulation," in IEEE Access, vol. 7, 2019, pp. 64837-64843, doi: 10.1109/ACCESS.2019.2914186.

[12] W. Hsu, D. Yaung, F. Hung, and K. Chou, "Phase Detection Autofocus Techniques, " US.S. Patent, 9804357B2, Oct. 31, 2017.

[13] C. Chan, S. Huang, and H. Chen, "Enhancement of Phase Detection for Autofocus," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 2017, pp. 41-45, doi: 10.1109/ICIP.2017.8296239.

[14] C. Ho, C. Chan and H. Chen, "AF-Net: A Convolutional Neural Network Approach to Phase Detection Autofocus," in IEEE Transactions on Image Processing, vol. 29, pp. 6386-6395, 2020, doi: 10.1109/TIP.2019.2947349.

[15] X. Zhang, Z. Liu, M. Jiang, M. Chang, "Fast and Accurate Autofocusing Algorithm Based on the Combination of Depth from Focus and Improved Depth from Defocus," Optics Express, vol. 22, no. 25, 2014, pp. 31237-31247.

[16] S. Yasugi, Khang Nguyen, K. Ezawa and T. Kawamura, "Depth from Defocus applied to Auto Focus," 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE), Tokyo, Japan, 2014, pp. 171-173, doi: 10.1109/GCCE.2014.7031237.

[17] H. Tang, S. Cohen, B. Price, S. Schiller, K. N. Kutulakos, "Depth from Defocus in the Wild," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 4773-4781, doi: 10.1109/CVPR.2017.507.

[18] C. Herrmann, R. Bown, N. Wadhwa, R. Garg, Q. He, J. Barron, R. Zabih, "Learning to Autofocus," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 2230-2239.

[19] C. Wang, Q. Huang, M. Cheng, Z. Ma and D. J. Brady, "Deep Learning for Camera Autofocus," in IEEE Transactions on Computational Imaging, vol. 7, 2021, pp. 258-271, doi: 10.1109/TCI.2021.3059497.

[20] A. Anikina, O. Rogov and D. Dylov, "Detect to Focus: Latent-Space Autofocusing System With Decentralized Hierarchical Multi-Agent Reinforcement Learning," in IEEE Access, vol. 11, pp. 85214-85223, 2023, doi: 10.1109/ACCESS.2023.3303844.

[21] N. Kehtarnavaz and H.-J Oh, "Development and Real-Time Implementation of a Rule-based Auto-Focus Algorithm," Real-Time Imaging, vol. 9, no. 3, 2003, pp. 197-203, doi: 10.1016/S1077-2014(03)00037-8.

[22] J. He, R. Zhou and Z. Hong, "Modified Fast Climbing Search Autofocus Algorithm with Adaptive Step Size Searching Technique for Digital Camera," IEEE Transactions on Consumer Electronics, vol. 49, no. 2, 2003, pp. 257-262, doi: 10.1109/TCE.2003.1209511.

[23] C. Park, S. Cha, Y. Lee, O. Kwon, D. Park, K. Kwon, J. Lee, "A highly accurate piezoelectric actuator driver IC for auto-focus in camera module of mobile phone," Proceedings of 2010 IEEE International Symposium on Circuits and Systems, Paris, France, 2010, pp. 1284-1287, doi: 10.1109/ISCAS.2010.5537267.

[24] C. L. Hsieh and C. S. Liu, "Design of a Voice Coil Motor Actuator With L-Shape Coils for Optical Zooming Smartphone Cameras," in IEEE Access, vol. 8, pp. 20884-20891, 2020, doi: 10.1109/ACCESS.2020.2968723

[25] Z. Wang, M. Lei, B. Yao, Y. Cai, Y. Liang, Y. Yang, Xibin Yang, Hui Li, and Daxi Xiong, "Compact Multi-band Fluorescent Microscope with an Electrically Tunable lens for Autofocusing," Biomedical Optics Express, Optica, vol. 6, 2015, pp. 4353-4364, doi: 10.1364/BOE.6.004353/

[26] H. Oku and M. Ishikawa, "High-speed liquid lens for computer vision," 2010 IEEE International Conference on Robotics and Automation, Anchorage, AK, USA, 2010, pp. 2643-2648, doi: 10.1109/ROBOT.2010.5509471.

[27] S. W. Hasinoff, K. N. Kutulakos, F. Durand and W. T. Freeman, "Time-constrained photography," 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 2009, pp. 333-340, doi: 10.1109/ICCV.2009.5459269.

[28] D. Vaquero, N. Gelfand, M. Tico, K. Pulli and M. Turk, "Generalized autofocus," 2011 IEEE Workshop on Applications of Computer Vision (WACV), Kona, HI, USA, 2011, pp. 511-518, doi: 10.1109/WACV.2011.5711547.

[29] K. Yamato, H. Chiba and H. Oku, "High Speed Three Dimensional Tracking of Swimming Cell by Synchronous Modulation Between TeCE Camera and TAG Lens," in IEEE Robotics and Automation Letters, vol. 5, no. 2, pp. 1907-1914, April 2020, doi: 10.1109/LRA.2020.2969911.

[30] 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," 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 2022, pp. 6637-6642, doi: 10.1109/IROS47612.2022.9981485.

[31] P. Grossmann, "Depth From Focus," Pattern Recognition Letters, ScienceDirect, vol. 5, no. 1, 1987, pp. 63-69.

[32] J. Martel, L. Muller, S. Carey, J. Muller, "Real-time Depth From Focus on a Programmable Focal Plane Processor," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 65, no. 3, 2018, pp. 925-934.

[33] J. Canny, "A Computational Approach to Edge Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 6, pp. 679-698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.

[34] Jaided AI, "EasyOCR," Ver. 1.7.1, 2023, https://github.com/JaidedAI/EasyOCR.

[35] E. Canas, J. Domanski, R. Kurmann, M. B. Currie, "QReader," Ver. 3.14, 2024, https://github.com/Eric-Canas/qreader.

[36] OpenCV, "Open Source Computer Vision Library (OpenCV)", Ver. 4.10.0.84, https://opencv.org/.

[37] Glenn Jocher and Ayush Chaurasia and Jing Qiu, "Ultralytics YOLOv8," Ver. 8.0.0, 2023, https://github.com/ultralytics/ultralytics, AGPL-3.0.

[38] A. Lukesic, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative Correlation Filter with Channel and Spatial Reliability," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 4847-4856, doi: 10.1109/CVPR.2017.515.