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

FoveaCam++: Systems-Level Advances for Long Range Multi-Object High-Resolution Tracking

Yuxuan Zhang[1] and Sanjeev J. Koppal[2]

***Abstract*— UAVs and other fast moving robots often need to keep track of distant objects. Conventional zoom cameras commit to a particular viewpoint, and carrying multiple zoom cameras for multi-object tracking is not feasible for power limited robotic systems. We present a dual camera setup that allows tracking of multiple targets at nearly 1km distance with high-resolution. Our setup includes a wide angle camera providing a conventional resolution view and a MEMS driven zoom camera that can query a specific region within the wide angle camera (WAC). We built and calibrated the two-camera system and implemented a real-time image fusion pipeline. We show multi-object tracking and stabilization in real world scenarios.**

I. Introduction

Many autonomous robots operate in environments where objects of interest are *few* and *far between*. For example, conventional cameras taken from fast moving aerial robot or water-surface vehicles produce images where most of the pixels are either sky or sea. The objects of interest are usually at the viewing limit, subtending to just a few pixels within the massive camera sensor's megapixel resolution.

The idea of actively zooming into multiple targets has been studied in robotics and active vision [1], [2], but these have been constrained by mechanical pan-tilt-zoom (PTZ) cameras. In contrast, recent work has taken inspiration from foveation in biology [3], [4], and created microelectromechanical (MEMS) based cameras that distribute resolution on areas of interest with a tiny scanning mirror. These devices can provide faster imaging than PTZ and can image multiple targets nearly simultaneously.

In this paper, we present **system-level advances** that enable the next generation of foveated camera which we term as FoveaCam++. Our system consists of a zoom lens, MEMS mirror, wide angle camera along with an embedded computer system for real-time performance. This system can be mounted onto a medium or heavy lift drone or similar robotics platform. Our system has approximately 1kg net weight and occupies a 20cm cubic volume. Compared to the previously available foveated cameras [5], [6], our system has the following advantages:

1) Longer range compact zoom lens: We have integrated a compact variable zoom lens with a MEMS mirror that allows programmable zoom for multiple ROIs and can image multiple objects from up to $1km$ distance*.
2) Adaptive depth-invariant calibration: We show a calibration of the zoom lens and the wide angle camera that enables in-situ adjustment during capture to remove parallax.
3) Multi-threaded software architecture: Our custom developed software stack optimizes frame rate, data throughput, robustness and latency.
4) Applications such as tracking and stabilization: We demonstrate applications for robotic platforms, such as a stabilization and tracking, in real-time for multiple objects, demonstrating increased resolution compared to conventional imaging.

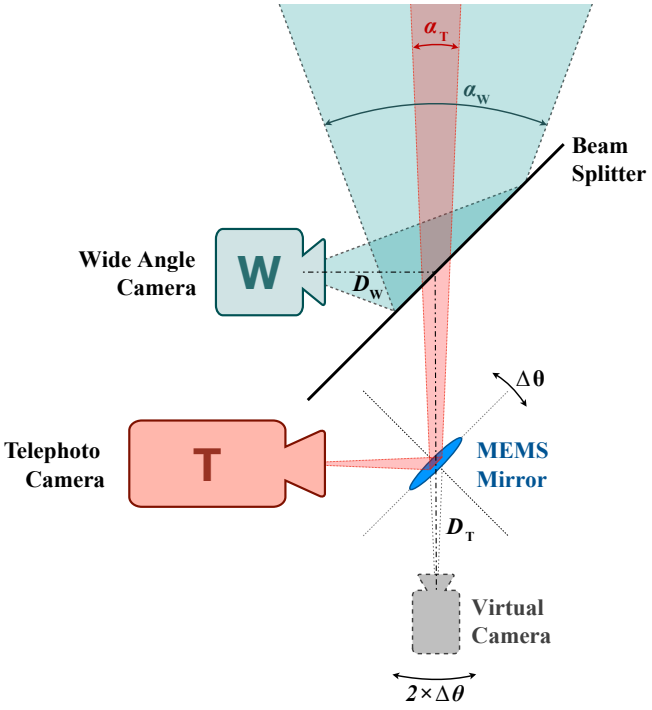


Fig. 1: Demonstrative diagram of the dual camera layout.

A. Previous Work

MEMS-based adaptive optics: Foveated camera designs (cameras capable to move their FOVs programmatically) have been used to mimic biological vision, with a low-res wide angle camera and a high-res zoom camera, have

[1,2] Yuxuan Zhang (zhangyuxuan@ufl.edu) and Sanjeev J. Koppal are with the Department of Electrical and Computer Engineering, University of Florida. Sanjeev J. Koppal holds concurrent appointments as an Associate Professor of ECE at the University of Florida and as an Amazon Scholar. This paper describes work performed at the University of Florida and is not associated with Amazon.

*Estimated based on real world experiments, varies on size of target and desired definition.

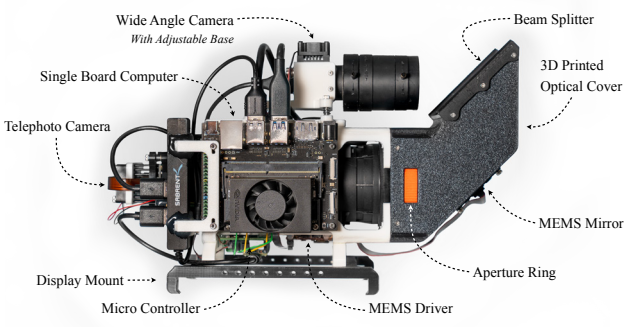


Fig. 2: FoveaCam++ Prototype

been proposed for static scenes, such as [7], [8], [9]. Tilmon et al. extended a MEMS mirror based system to moving scenes with up to two targets at $5m$ distance [6]. Compared to previous efforts, ours is the first that can track multi-targets at up to $1km$ distances. All these MEMS-modulated imaging techniques are tangentially related to the adaptive removal of atmospheric turbulence effects [10]. Our use of MEMS mirror enables advantages in compactness, high-speed and low wear-and-tear compared to mechanical PTZ methods [11].

Active vision and adaptive sampling: This paper is about system-level advances for a new type of long range foveated camera. The idea of paying attention to ROIs has been long studied in attention, active vision and robotics [12], [13], [14], [15], [16], [17], [18], [19], [20]. Robotic and active vision-based algorithms could be implemented on this platform, and our work complements these methods.

MEMS/Galvo mirrors for vision and graphics: MEMS and Galvo mirrors are mostly used for modulating light in projectors and lasers for applications in vision and graphics [21], [22], [23], [24], [25], [26]. In our work, we used a reflective mechanism (beam splitter) to avoid restriction of bandwidths for the passive sensor. This setup makes it possible to incorporate different types of wide angle sensors, such as IR or event camera, for different use cases. In addition, tracking with large galvo mirrors has been shown in [27], [28]. These devices usually require two mirrors for modulating both azimuth and elevation while a single MEMS can rotate in two dimensions, contributing to improved compactness and robustness of our proposed system.

II. Design

Conventional zoom entails a trade-off between angular resolution and field-of-view (FOV). For a system with focal length f, sensor resolution N, pixel pitch d_{px} and FOV ω_{fov}, the angular resolution can be given by $\frac{f \cdot tan(\omega_{fov})}{\omega_{fov} \cdot d_{\text{px}}}$, and as the FOV approaches zero (i.e. near the center of the sensor), we can simplify this as $\frac{f}{d_{\text{px}}}$ [†]. On the other hand, the FOV itself reduces by the ratio $\arctan\left(\frac{f}{N \cdot d_{\text{px}}}\right)$, i.e. the system loses details as its FOV enlarges, or get a narrower FOV as it zooms into a smaller feature.

[†]This equation gives $^{\text{px}}/_{\text{rad}}$. Multiply by $\frac{\pi}{180}$ to convert to $^{\text{px}}/_{\text{deg}}$.

Our idea is to break these limits by reflecting the zoom camera off a microelectromechanical (MEMS) mirror, whose scanning pattern is controlled. In this sense, the zoom camera is not committed to one region and can be quickly swiveled to different regions. This foveated camera design [6] works because the azimuth and elevation are controlled by voltages over time, $(\theta(V(t)), \phi(V(t))$, over the mirror FOV ω_{mirror}.

A. Hardware Design for Robotic Applications

Although [6] was able to show feasibility over short distances for two targets, our approach solves multiple novel system level problems to make FoveaCam++, which is better suited to robotic applications such as imaging from drones.

Hardware Prototype Summary: The complete assembly of the camera system implementing Fig.1 is shown in Fig.2. As is shown in the figure, a wide angle camera is mounted on the front side of the assembly. It is directly pointed to a beam-splitter mounted on the black 3D printed cover. The lens installed on this wide angle camera was configured to cover the full foveated field of view of the telephoto camera. Behind the wide camera is a Kurokesu motorized zoom lens with a FLIR board level sensor attached to it. The sensor can run at up to 226fps given enough light. Both cameras are connected to a single board computer sitting on the top of the assembly via USB3, as is represented by dashed lines of the block diagram shown in Fig.3. The single board computer is in charge of coordinating all components in the system. A micro-controller and a MEMS voltage regulator are mounted to the back side of the camera (behind the display mount). The 3D printed structure holding the MEMS mirror is designed to be modular so it can be swappable without disassembling the entire device.

Summary of Improvements: Previous efforts in foveated cameras had avoided color imagery, since inter-reflections in the MEMS packaging and cover glass would cause ghosting [6]. We obtained an updated MEMS package design from the supplier and developed custom 3D printed structures to deal with this ghosting issue. We also have included a secondary wide angle camera coupled with a beam-splitter to the MEMS-modulated zoom camera. The wide angle camera will provide awareness to broader scene which can be used to help the foveated camera to determine an ROI with interested targets. We have implemented a closed loop calibration between these two cameras, which can overcome sources of noise such as thermal expansion/contraction or mechanical displacement. Finally, the latency of data processing determines how fast FoveaCam++ can respond to moving targets. To enable real-time speed, we use multi-threaded software to process high volume of data in parallel.

B. Mirror Control and Synchronization

Key to the high quality results in this paper are synchronization advances as shown in Fig.3. In previous work[6], the voltage regulator V that controls the MEMS mirror was directly connected to a single board computer's GPIO pins. It lacks a way to synchronize the mirror with the foveated sensor. Consequently, the system relies on USB data

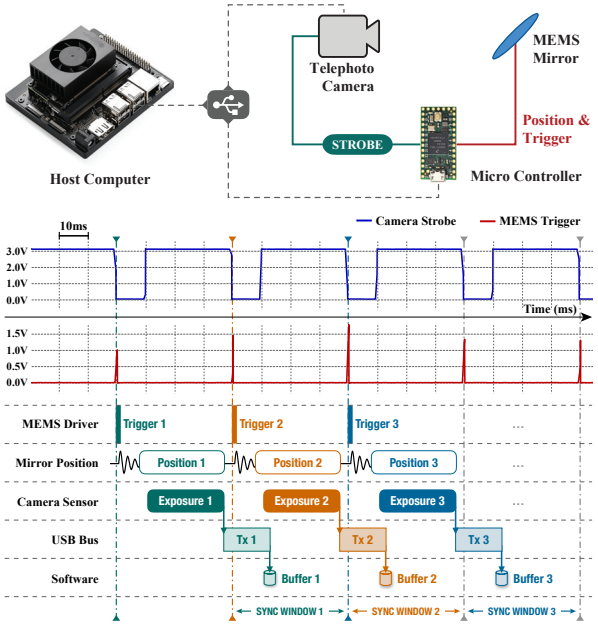


Fig. 3: Upper: block diagram of camera-mirror synchronization; Lower: signal waveform v.s. sync timing chart.

transmission to determine when to move the mirror to next position. This method lacks consistency of frame-position correlation and causes image quality issues such as double exposure (ghosting).

In our work, we addressed this challenge by introducing a strobe synchronization mechanism. A micro-controller unit (MCU) was used to handle the strobe synchronization signal from the foveated sensor. Instead of directly send the MEMS mirror position to the mirror, the MCU will hold the next position sent from host computer until next falling edge of the sensor strobe signal. Our custom developed embedded firmware for the MCU provides near real-time response to the strobe synchronization signal. Such solution enables very tight synchronization and therefore allows much higher frame rate without risking to break the mirror-sensor synchronization. A waveform of camera strobe sync signal v.s. MEMS trigger signal is plotted in Fig.3, showing an observed delay of less than 200ns.

C. Software Architecture

A multi-threaded software architecture was designed to handle large amount of data throughput generated by the dual-camera setup while minimizing the delay of data processing. As is shown in Fig.4, the system was divided into modules which runs in separate threads. Common components such as camera capturing thread and MEMS mirror control/synchronization threads are implemented as separate threads. In addition to these common threads, a set of data pipes are exposed to each task. These data pipelines are designed to minimize inter-thread communication latency and maximize data throughput. These pipelines define a standard way to communicate with the underlying hardware, abstracting away most of the complexity (such as communication protocols, recovery from erroneous states,

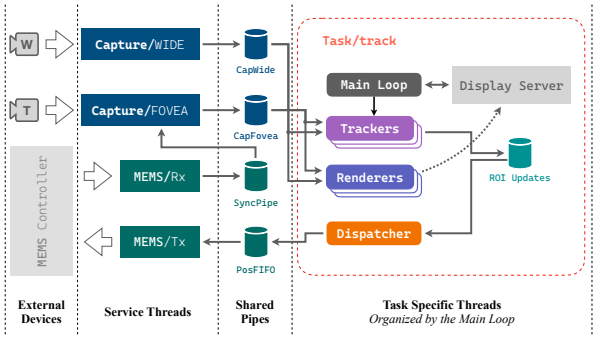


Fig. 4: Diagram of the multi threaded software architecture. Using task "track" as an example.

Type of Camera	Field of View		Pixel Pitch	Focal Length	Angular Resolution[a]
	Horizontal	*Vertical*			
Wide Angle	14.26°	10.69°	3.75 µm	19.8 mm	92.24 px/deg
Foveated	1.06°	0.79°	3.75 µm	267.5 mm	1245.13 px/deg
Comparison[b]	26.32°	17.11°	1.00 µm	9.0 mm	157.08 px/deg

[a] *Angular resolution is calculated according to equations given in* Sec.II
[b] *Using the telephoto camera on iPhone 13 Pro for comparison.*

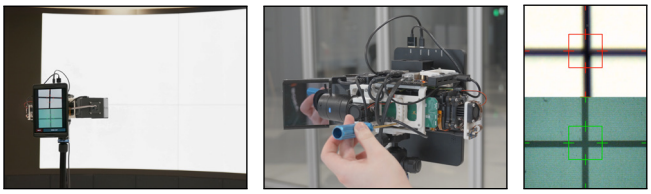
TABLE I: Key parameters of cameras in the system

congestion resolution, etc.) when implementing a specific workflow ("task").

"Tasks" are defined as a set of UI layouts and underlying logic and algorithm modules that, when working together, can serve a specific purpose. For example, the "track" task demonstrated in Fig.4 contains 3 tracker threads and 6 render threads. The main loop processes user input and translates them into ROIs for tracker threads to start with. The tracker threads wait for initial ROIs from the main loop and continuously update their ROIs using frames from "CapWide" and "CapFovea" data pipes. The trackers will post their latest tracking ROIs to another data pipe "ROI Updates", which is monitored by the dispatcher thread. The dispatcher thread will dispatch the MEMS mirror to each ROI in an interleaving pattern. The resulting frames will then be rendered into their corresponding tiles on display.

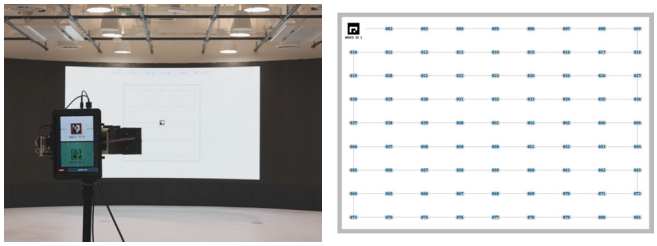
D. Obtained FOV and Angular Resolution

In our setup, the telephoto view is capable of foveating within a rectangular area constrained by the MEMS mirror. A small number of margin pixels are reserved on the wide angle view to account for possible calibration drifts as discussed in Sec.III-B. Parameters of both of the cameras have been measured during the first calibration step and are listed in Table.I. It outlines the superiority of angular resolution achieved by our setup. The foveated camera has an angular resolution 13.6 times higher than its wide angle counterpart (they have identical CMOS sensor), or 8.9 times higher than iPhone 13 Pro's telephoto camera. In addition, the foveated camera has extra degrees of freedom, giving it an extended field of view. The MEMS mirror has an tilting range of 10.63° horizontally and 7.92° vertically. Using the numbers



(a) (b) (c)

Fig. 5: (a) Setup of dual camera alignment, camera was pointed to a large LED panel showing crosshair; (b) WAC being adjusted to align with the telephoto camera; (c) Wide angle view (top) and telephoto view (bottom) after alignment.



(a) (b) (c) (d)

Fig. 6: (a) Setup of mirror calibration, camera was pointed to the same LED panel as the alignment process but moved further away; (b) The calibration utility automatically generates calibration points; (c) WAC view, sample points already captured are shown with blue markers; (d) Telephoto view showing the detection of the last sample point (ID 81).

extracted from the calibration step, we can find the extended field of view to be $11.69°$ by $8.71°$.

III. Alignment and Calibration

We have two cameras, a high-resolution foveated sensor that images reflections off the MEMS mirror, and a low-resolution wide angle camera. Using the center axis of the WAC as the origin, its projection matrix can be defined as $\mathbf{K_1}\left[\ \mathbf{I}\ |\ \mathbf{0}\ \right]$, where $\mathbf{K_1}$ is the intrinsic matrix. We calibrate this camera by recording a sequence of checkerboard patterns and performing camera intrinsic extraction.

Let the second camera's projection matrix be $\mathbf{K_2}\left[\ \mathbf{R}\ |\ t\ \right]$, where $\mathbf{K_2}$ is the intrinsic matrix and R and t are the rotation and translations between the center of projections of the two cameras. Our design strategy is (1) align the internal camera parameters to maximize similarity between their intrinsic matrices, i.e. $\mathbf{K_1} \approx \mathbf{K_2}$ and (2) use a beamsplitter such that, for scenes at a large distance Z, the translation can be ignored, i.e. $|t| << Z$. With such design, calibration reduces to just estimating a rotation $\mathbf{R}$. For mechanical setups with

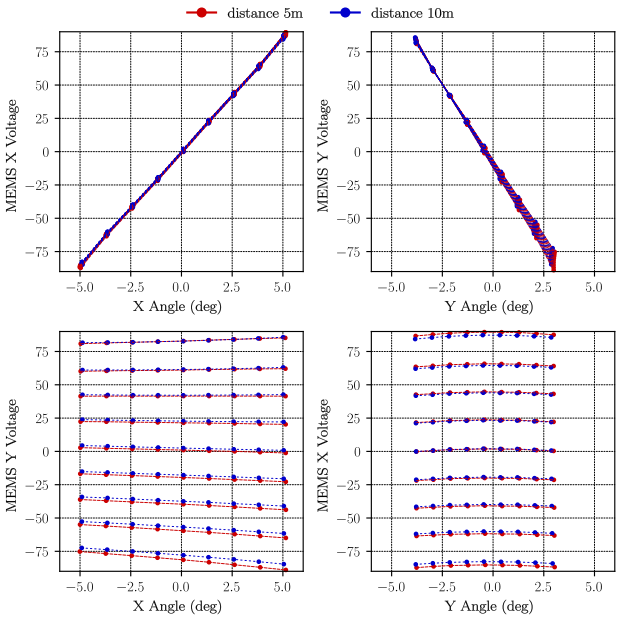


Fig. 7: Calibration points captured at different distances.

aligned roll angles, the rotation becomes two angles of a distant appearance field defined on the hemisphere of directions.

A. Dual Camera Alignment and Coverage Adjustment

We first match the coverage of view from the wide angle camera to the telephoto camera. This is achieved by centering a cross-hair marker on both views, as shown in Fig.5. Adjustments are done through a specially designed, fine-tunable spring loaded base shown in Fig.5b. Another tool was developed and used to perform a straightforward "coverage" check to make sure that the FOV of the foveated camera is fully contained in the wide angle view. "Full coverage" means for any possible mirror location, the content seen in the foveated view will also be fully covered in the wide angle view. The wide angel lens will be adjusted if the coverage is either too wide or too narrow.

B. Mirror Calibration and the Regression Model

With the two camera aligned and coverage adjusted, we can calibrate the rotation between the two cameras. The objective of this process is to obtain a mapping from an ROI of the wide angle camera to the MEMS mirror's input voltages and vice versa. To obtain this mapping, we image a moving ArUco marker on a large screen. We track the movement of the marker using a PID controller so the marker stays at the center of the foveated view. The MEMS mirror voltage of each sample point is recorded against the pixel coordinate of the center of marker in wide angle view.

The test points captured in Sec.III-B were plotted in Fig.7. Mapping between MEMS mirror voltages and wide angle camera's pixel offset can be derived from these points using linear regression. In order to closely match the test points, we introduced a cross correlation term in the regression matrix. As is described in Eq.1, the linear regression model is able

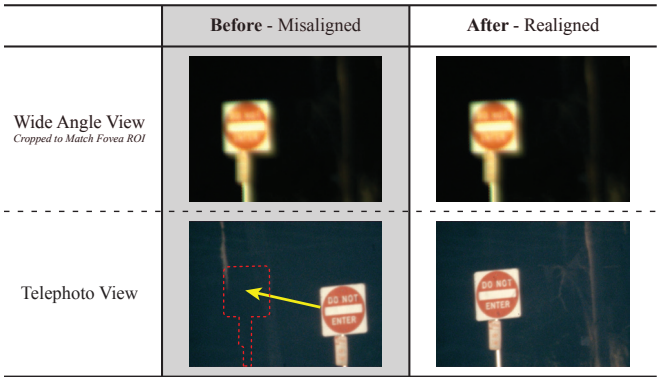


Fig. 8: Wide angle and telephoto views before and after incremental drift compensation.

to correlate from a given set of voltage values to a pixel location on the wide angle view and vice versa.

$$\begin{aligned} k_i &= \begin{bmatrix} a_i & b_i & c_i & d_i \end{bmatrix} \\ V_x &= k_1 \cdot \begin{bmatrix} P_x & P_y & (P_x \cdot P_y) & 1 \end{bmatrix}^T \\ V_y &= k_2 \cdot \begin{bmatrix} P_x & P_y & (P_x \cdot P_y) & 1 \end{bmatrix}^T \\ P_x &= k_3 \cdot \begin{bmatrix} V_x & V_y & (V_x \cdot V_y) & 1 \end{bmatrix}^T \\ P_y &= k_4 \cdot \begin{bmatrix} V_x & V_y & (V_x \cdot V_y) & 1 \end{bmatrix}^T \end{aligned} \quad (1)$$

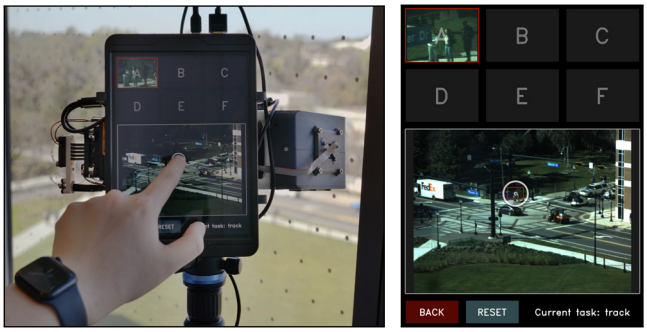
As a final step, we incorporated automatic incremental adjustment to account for calibration drift during field application. Such drift could be caused by multiple factors such as thermal inflation/contraction of 3D printed parts. In the "match" task of the user interface, one can manually point the foveated camera to a feature-rich region of the scene. Then, the user can click "calibrate" button to start this process. Cross-correlation is used to obtain relative drift of the original calibration. The drift vector will then be fed back to the regression model. As is shown in Fig.8, this approach was able to cancel out most of the drift.

IV. Field Experimentation Results

A. Multi-Target Tracking

The previous section demonstrated the calibration process to create a mapping between pixel location and MEMS position. Given this mapping, wide angle camera can be used to provide tracking information for the foveated camera. The foveated telephoto camera will be used to capture smooth high-res video of multiple tracked objects. The objects to track can be initialized either by an automatic detection (e.g. face detection) or manually through a touch UI (as shown in Fig.9). For each target, a kernalized correlation filter (KCF) object tracker [29] will be dispatched as an individual thread to consciously update tracking ROI. And the MEMS controller will interleave across each target according to their latest ROIs. Frames for each target are then multiplexed into their corresponding tiles and rendered to the screen.

As is presented in Fig.10, samples are taken from multiplexed video streams, each multiplexed view receives a live stream. Wide angle view is shown on the 1st column,



(a) (b)

Fig. 9: (a) Interactive tracking GUI; (b) Screenshot;

with each ROI annotated and labeled with numbers. The corresponding foveated view are shown in the 2nd column and an ROI matched counterpart cropped from wide angle view is supplied on the 3rd column.

The foveated view T2 of Fig.10 (c) clearly reveals the superiority of angular resolution of the foveated camera (covered in Sec.II-D). For example, the text displayed on the bus front signage was clear to read from the foveated telephoto view, but is not directly readable from the cropped wide angle view.

B. Stabilization

Our system is also capable of simulating a stabilized gimbal with extended range of motion. In this scenario, we disengage the synchronization between MEMS mirror and camera strobe in order to continuously update the position of MEMS mirror **even when an image is being exposed on the foveated camera sensor**. In other words, we expect to reduce motion blur relative to the target.

We implemented a linear predictive motion model to compensate for data transmission delay, and smooth the MEMS mirror motion independent of the sample rate of the wide angle camera. The algorithm computes the transient velocity $\boldsymbol{v}(t_N)$ by dividing the transient spacial offset Δx with time increment Δt. Additionally, a temporal convolution described by Eq.2 was incorporated to mitigate noises introduced by the tracking kernel. The "decay" parameter k balances the trade off between motion smoothness verses temporal responsiveness. As in Eq.3, this model performs an iterative weighted-add optimization without storing an array of data points.

$$\boldsymbol{v}_{\text{est}}(t_N) = (1-\boldsymbol{k}) \cdot \sum_{n=0}^{N} \boldsymbol{k}^{(N-n)} \cdot \boldsymbol{v}_{\text{obs}}(t_n) \quad (2)$$

$$with\ 0 < \boldsymbol{k} < 1\ and\ N = \frac{t_N - t_0}{\Delta t}$$

$$= (1-\boldsymbol{k}) \cdot \boldsymbol{v}_{\text{obs}}(t_N) + \boldsymbol{k} \cdot \boldsymbol{v}_{\text{est}}(t_{N-1}) \quad (3)$$

A configurable parameter t_{delay} is introduced in order to compensate for delays introduced in multiple processes,

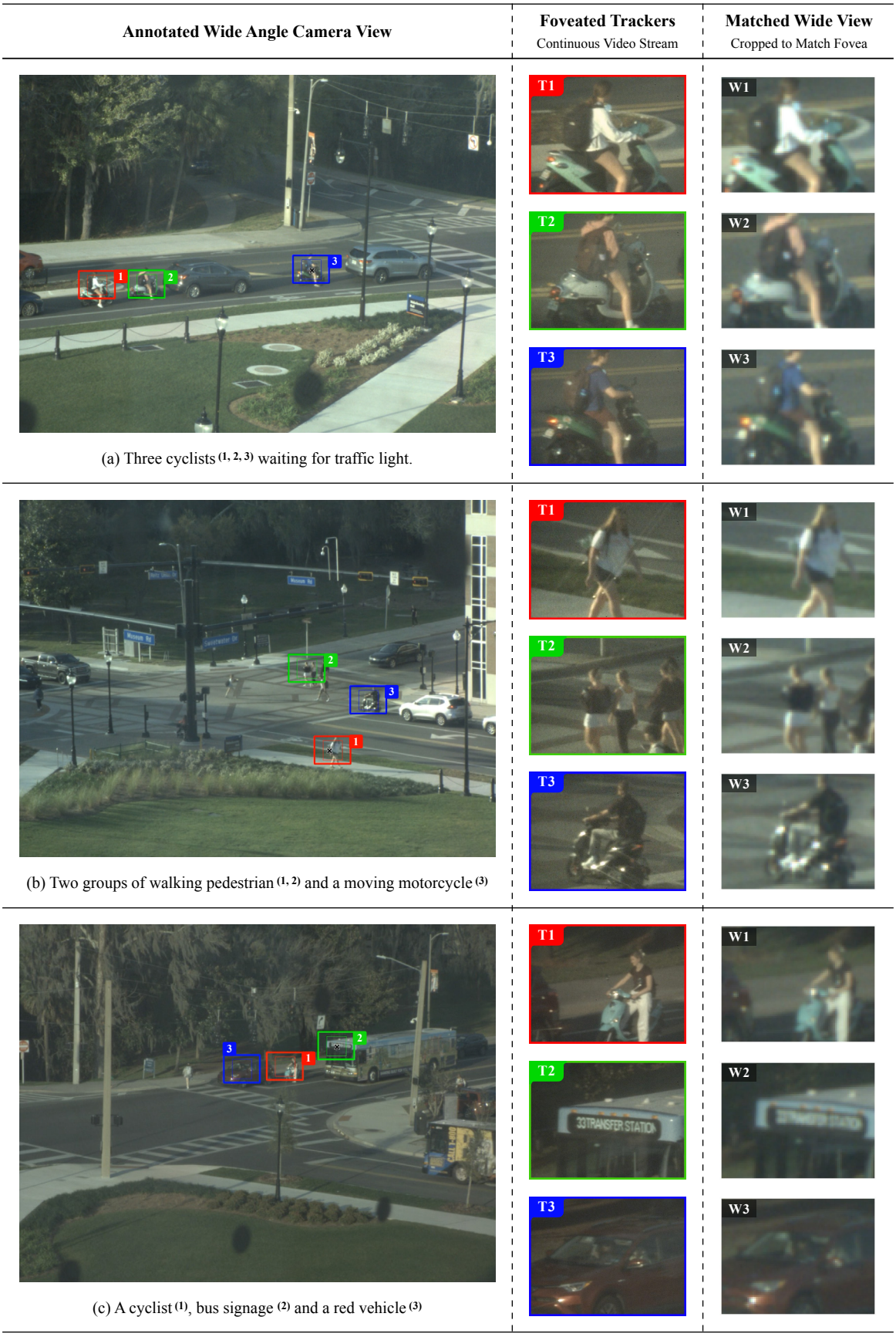
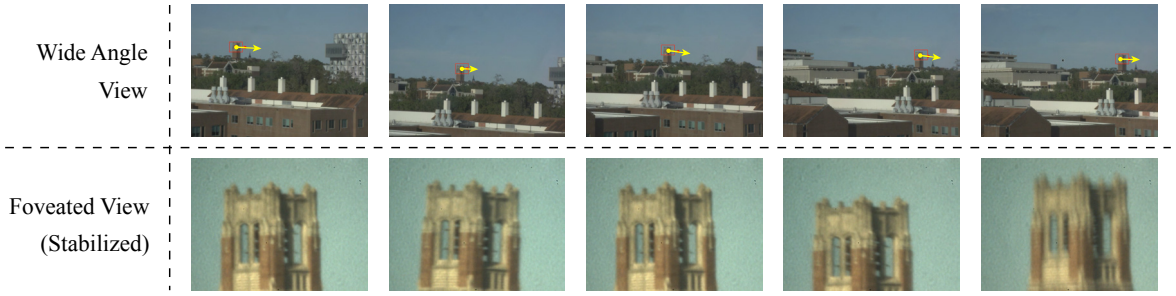
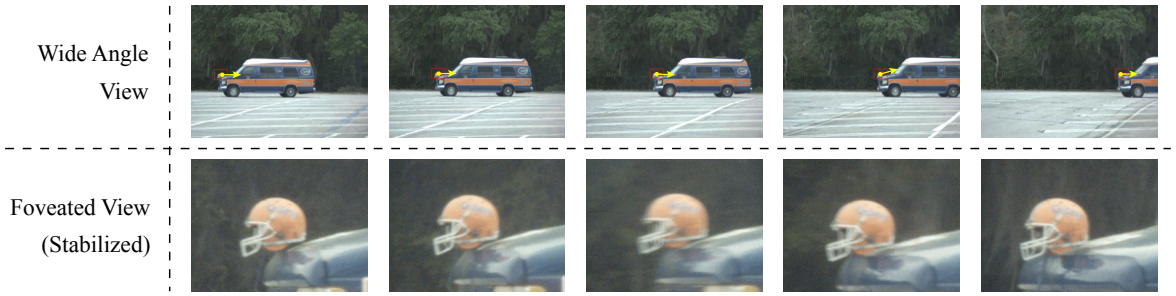


Fig. 10: Results of multi-object foveated tracking experiment.

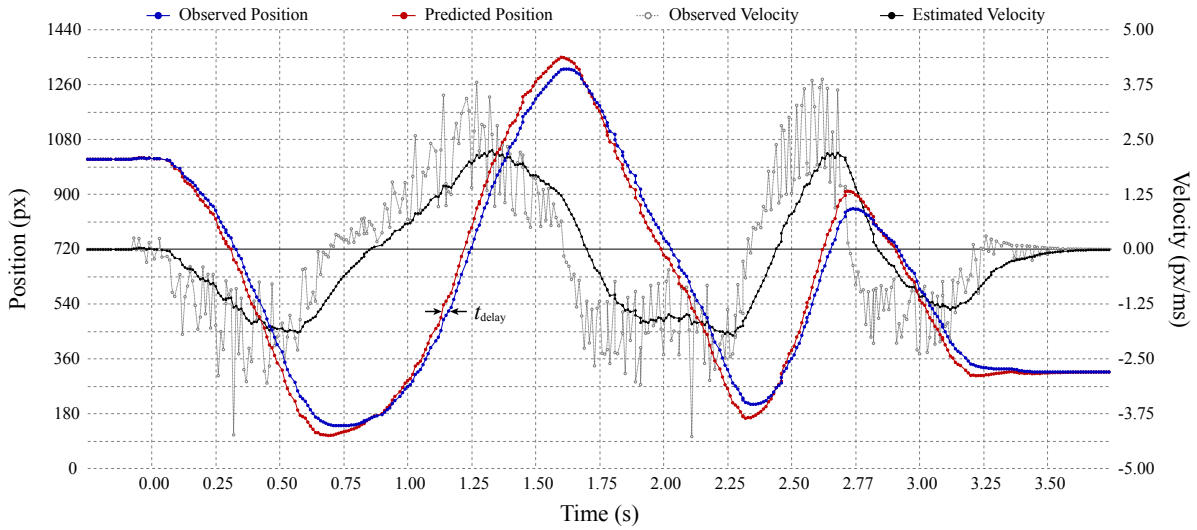


(a) Stabilization on a bell tower with the camera being swept manually. The top of the bell tower was kept at the center of telephoto camera's FOV all the time.



(b) Stabilization on a football helmet with the camera moving with a vehicle. Perspective transitions are visible in foveated view with very little translation

(c) Stabilization on a traffic sign with the camera mounted on a vehicle driving towards it. The sign stays in the center of fovea FOV even when driving over speed bumps.



(d) Data recorded when sweeping camera around a crosshair target, demonstrating the predictive algorithm.

Fig. 11: Results and transient data of foveated stabilization experiment.

including receiving frame buffers from camera sensors, and applying MEMS mirror's voltages. With that in consideration, we are able to derive the desired MEMS mirror position for any given time $x_N(\Delta t)$ as shown in Eq.4.

$$x_N(\Delta t) = x_N^0 + \boldsymbol{v}_{\text{est}}(t_N)\left(\Delta t + t_{\text{delay}}\right) \tag{4}$$

After establishing a motion prediction model, the algorithm continuously updates mirror position based on its own prediction. The update rate is substantially higher than the actual frame rate of both camera sensors, this helps to cancel the motion blur during each exposure. As shown in Fig.11d, transient tracking trajectories were recorded in a laboratory environment. The predicted position (red) stably maintains an offset of t_{delay} ahead of observed position (blue). In real world experiments shown in Fig.11, this algorithm kept the subject stably in its field of view while the system was moved either by hand or with a vehicle.

V. Conclusions, Limitations and Future Work

In this work, we presented the design and implementation of a high-performance foveated camera system. We developed a robust calibration process and proposed a feasible solution for calibration drift in the field, allowing the system to be effectively utilized outside of a laboratory environment. Building on this foundation, we demonstrated tracking and stabilization applications with performance that, to the best of our knowledge, surpasses any alternative systems of similar size and cost. Looking forward, we aim to continue refining the system and integrating it with robotic platforms to address real-world challenges that could benefit from the unique capabilities of FoveaCam++.

ACKNOWLEDGMENT

The authors wish to thank and acknowledge partial support from the ONR from grants N00014-18-1-2663 and N00014-23-1-2429 as well as the NSF from grants 1942444 and 2330416.

References

[1] J. Aloimonos, I. Weiss, and A. Bandyopadhyay, "Active vision," *International journal of computer vision*, vol. 1, no. 4, pp. 333–356, 1988.

[2] S. Chen, Y. Li, and N. M. Kwok, "Active vision in robotic systems: A survey of recent developments," *The International Journal of Robotics Research*, vol. 30, no. 11, pp. 1343–1377, 2011.

[3] S. Frintrop, E. Rome, and H. I. Christensen, "Computational visual attention systems and their cognitive foundations: A survey," *ACM Trans. Appl. Percept.*, vol. 7, no. 1, jan 2010. [Online]. Available: https://doi.org/10.1145/1658349.1658355

[4] J. M. Findlay and I. D. Gilchrist, *Active vision: The psychology of looking and seeing*. Oxford University Press, 2003, no. 37.

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

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

[7] G. Sandini and G. Metta, "Retina-like sensors: motivations, technology and applications," in *Sensors and sensing in biology and engineering*. Springer, 2003, pp. 251–262.

[8] S. Liu, C. Pansing, and H. Hua, "Design of a foveated imaging system using a two-axis mems mirror," in *International Optical Design Conference 2006*, vol. 6342. International Society for Optics and Photonics, 2006, p. 63422W.

[9] H. Hua and S. Liu, "Dual-sensor foveated imaging system," *Applied optics*, vol. 47, no. 3, pp. 317–327, 2008.

[10] J. M. Beckers, "Adaptive optics for astronomy: principles, performance, and applications," *Annual review of astronomy and astrophysics*, vol. 31, no. 1, pp. 13–62, 1993.

[11] T. Nakao and A. Kashitani, "Panoramic camera using a mirror rotation mechanism and a fast image mosaicing," in *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, vol. 2. IEEE, 2001, pp. 1045–1048.

[12] S. Frintrop, E. Rome, and H. I. Christensen, "Computational visual attention systems and their cognitive foundations: A survey," *ACM Transactions on Applied Perception (TAP)*, vol. 7, no. 1, p. 6, 2010.

[13] N. Bruce and J. Tsotsos, "Attention based on information maximization," *Journal of Vision*, vol. 7, no. 9, pp. 950–950, 2007.

[14] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT press, 2005.

[15] B. Charrow, G. Kahn, S. Patil, S. Liu, K. Goldberg, P. Abbeel, N. Michael, and V. Kumar, "Information-theoretic planning with trajectory optimization for dense 3d mapping." in *Robotics: Science and Systems*, vol. 11. Rome, 2015.

[16] E. Ristani and C. Tomasi, "Features for multi-target multi-camera tracking and re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6036–6046.

[17] T. Darrell, B. Moghaddam, and A. P. Pentland, "Active face tracking and pose estimation in an interactive room," in *Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 1996, pp. 67–72.

[18] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[19] H. Wei, P. Zhu, M. Liu, J. P. How, and S. Ferrari, "Automatic pan–tilt camera control for learning dirichlet process gaussian process (dpgp) mixture models of multiple moving targets," *IEEE Transactions on Automatic Control*, vol. 64, no. 1, pp. 159–173, 2018.

[20] C. Ding, B. Song, A. Morye, J. A. Farrell, and A. K. Roy-Chowdhury, "Collaborative sensing in a distributed ptz camera network," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3282–3295, 2012.

[21] T. P. Flatley, "Spacecube: A family of reconfigurable hybrid on-board science data processors," 2015.

[22] B. L. Stann, J. F. Dammann, M. Del Giorno, C. DiBerardino, M. M. Giza, M. A. Powers, and N. Uzunovic, "Integration and demonstration of mems–scanned ladar for robotic navigation," in *Proc. SPIE*, vol. 9084, 2014, p. 90840J.

[23] K. T. Krastev, H. W. Van Lierop, H. M. Soemers, R. H. M. Sanders, and A. J. M. Nellissen, "Mems scanning micromirror," Sep. 3 2013, uS Patent 8,526,089.

[24] T. Hawkins, P. Einarsson, and P. E. Debevec, "A dual light stage." *Rendering Techniques*, vol. 5, pp. 91–98, 2005.

[25] M. O ´ Toole, D. B. Lindell, and G. Wetzstein, "Confocal non-line-of-sight imaging based on the light-cone transform," *Nature*, vol. 555, no. 7696, p. 338, 2018.

[26] J. Wang, J. Bartels, W. Whittaker, A. C. Sankaranarayanan, and S. G. Narasimhan, "Programmable triangulation light curtains," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 19–34.

[27] 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 *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 6637–6642.

[28] T. Zhang, Z. Li, Q. Wang, K. Shimasaki, I. Ishii, and A. Namiki, "Dof-extended zoomed-in monitoring system with high-frame-rate focus stacking and high-speed pan-tilt adjustment," *IEEE Sensors Journal*, vol. 24, no. 5, pp. 6765–6776, 2024.

[29] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, p. 583–596, Mar. 2015. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2014.2345390