Utilizing Optically Multiplexed Cameras for Simultaneous Localization and Mapping (SLAM)

Tom Cheng
to26177@mit.edu

Ian Palmer
iapalm@mit.edu

Richard Showalter-Bucher
ri25917@ll.mit.edu

Abstract

Optically multiplexed (OM) cameras are a new technology that enables multi-camera performance from a single focal plane array. This technology is not widely adopted yet, so for our project we wanted to explore the feasibility of this technology in the area of simultaneous localization and mapping (SLAM). In our project, we developed the capability to generate synthetic OM imagery, explored techniques for mitigating structured-noise, and integrated various 2-field-of-view (2-FOV) OM monocular and OM stereo into an open-source SLAM architecture.

1. Introduction

In traditional cameras, every pixel maps to one area in the scene. Thus, they suffer from a field of view (FOV) and resolution trade-off. Increasing the FOV can only be done by either increasing the number of pixels, which is costly and results in higher data bandwidth, or increasing the FOV of the optics which results in reduced resolution and increased optical distortion. Current methods for increasing the effective FOV with traditional cameras include using multiple cameras or using a single camera with a scanning mechanism. They both suffer from increased cost and size, and the latter suffers from increased mechanical complexity and a low temporal sampling rate.

An optically multiplexed (OM) camera represents a paradigm shift in how wide-FOV images are captured – each pixel is used to capture multiple areas in the scene simultaneously (Figure 1). It overcomes the FOV and resolution trade-off, and results in minimal increases in size and weight to the imaging system.

There is increasing interest in enabling SLAM algorithms to run on miniaturized autonomous vehicles (AVs). They are motivated by the potential for autonomous vehicles to navigate in confined spaces, such as inside buildings or even the human body [6]. Small drones are also more affordable due to decreased use of materials and power of the motors, and – coupled with their increased portability – are more amenable to widespread use in civilian applications [9]. The small form factor of a miniaturized AV limits the amount of energy that can be stored for use by the AV's electronics [4], which in turn limits the weight of the vehicle because its motors must function within the limited energy budget. OM cameras address the attendant issues of miniaturizing AVs. They eliminate the need for two cameras in an AV that utilizes stereo imagery to navigate, thereby reducing both power and weight in the system. In AVs that use monocular imagery to navigate, an OM camera can increase their navigational accuracy by effectively increasing the monocular imaging system's FOV and providing more environmental features for the SLAM algorithm to process.

In addition, recent developments in OM camera technology have reduced their size to no more than the size of a single lens [11], meaning that there is little or no size and weight penalty to using OM cameras over traditional cameras. Given the potential advantages of integrating OM cameras into SLAM for AVs and their advantage over traditional cameras in size and energy-constrained platforms, we feel there is value in assessing the potential for SLAM algorithms to utilize OM cameras.

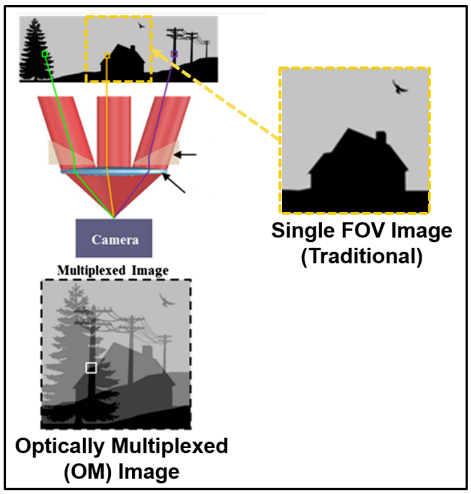

Figure 1. Optically Multiplexed Camera

2. Related Work

The primary application envisioned by early designers of OM imaging systems has been wide-FOV surveillance in security and commercial contexts [8]. More recent developments in OM camera design have drastically reduced the size and weight of these systems, through more efficient ways of optically multiplexing multiple FOVs, such as by dividing the aperture of a lens rather than using a complicated assembly of beamsplitters [11]. This has opened the possibility of introducing OM cameras to mobile applications that require low size, weight, power, and cost (SWaP-C), such as navigation and optical communication [8]. Current efforts in the OM imaging area have continued to focus on refining the optical design and devising improved ways of reconstructing the original, non-multiplexed scene from the OM image [7]. Less attention has been given to exploring the value and feasibility of replacing traditional cameras in mobile applications such as autonomous navigation.

3. Approach

3.1. OM Challenges

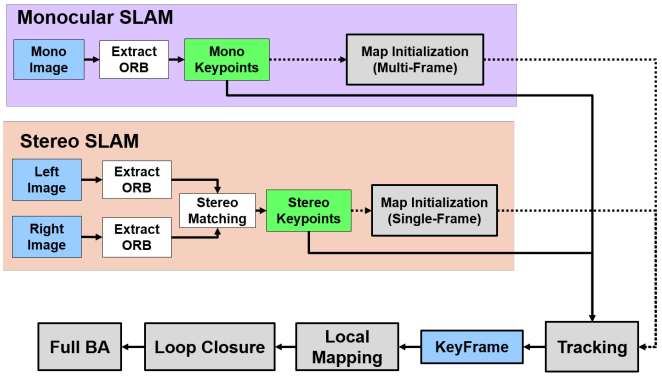

Figure 2. Overview of Orb-SLAM2 Architecture

Two challenges associated with integrating OM cameras into a SLAM architecture are FOV ambiguity and structured noise. Both of these are caused by the superposition of FOVs onto a single focal plane array as shown in Figure 3.

These challenges manifest as incorrect feature detections (from structured noise), feature matching between frames (from structured noise), and 3D point cloud generation (from FOV ambiguity) within SLAM.

For the project we sought to develop methods that mitigate these challenges in an open-source SLAM architecture, ORB-SLAM2. Figure 2 shows the architecture ORB-SLAM2 which can support both monocular and stereo imagery.

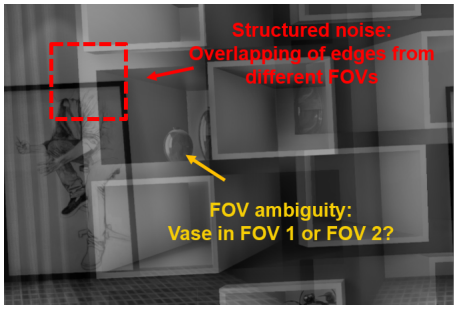

Figure 3. OM Camera Challenges

3.2. Synthetic Imagery Generation

We generated imagery for the project via the FlightGoggles framework. This framework was designed to render photorealistic hardware-in-the-loop drone imagery for testing SLAM performance in various synthetic environments [10]. Figure 4 shows some of the available scenes.

We wanted to use a realistic trajectory to feed into FlightGoggles, so we utilized the Blackbird dataset [5] of drone trajectories. In this paper we show the results for the Butterfly World scene and a constant-yaw ampersand-shaped trajectory.

Figure 4. Available environments in FlightGoggles (from top to bottom and left to right): Butterfly World, NYC Subway, Museum, and Hazelwood Loft. The environment outlined in red was used in our tests.

3.2.1 OM Camera Model

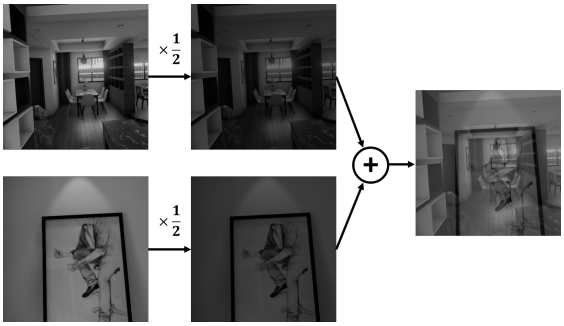

Figure 5. OM Model

We used a subdivision-of-aperture model to produce our OM images. This model divides the pixel intensities of each FOV by the number of FOVs being multiplexed, then adds them together. This follows the standard technique of models used in current OM research and development [3]. This model is shown Figure 5.

We also implemented the ability to horizontally flip one of the FOVs so that we could model non-reversing/reversing mirrors for our camera. These mirror models were used in our monocular OM epipolar constraint approach (Sect. 3.4.1).

3.2.2 Spectral Filtering (SF) Model

To produce a spectrally filtered image, we took the synthetic color imagery and attenuated each color channel by the convolution of normalized quantum efficiency (QE) curves and Cyan/Red color filter transmittance values. We used the QE of a Point Grey Flea-3 focal plane array [2] and color filter transmittances from [1]. We also applied a mosaic (BGGR) and demosaic to further degrade the image to model a realistic performance of a color camera. For the demosaicing, we applied a linear interpolation of the red and green channels, and an interpolation of the blue-green difference for the blue channel.

3.3. Feature Extractor Analysis

In the context of SLAM, we observed that the primary effect of the structured noise that comes from superimposing multiple non-overlapping FOVs is the generation of features (corners) that do not belong to any objects in the FOVs. These artifacts, or "false corners", commonly arise when edges from different FOVs with different luminances intersect in the multiplexed image, as shown in Figure 3.

State-of-the-art feature detectors such as SIFT, SURF, and ORB (used in ORB-SLAM2) measure relative pixel intensities in a small neighborhood to identify corners in an image [1]. We found, however, that these feature detectors are susceptible to producing false positives at false corners.

Because false corners do not belong to any objects in the FOVs, they do not stay fixed in position relative to any object, between consecutive frames. Thus, estimating the camera's trajectory from the movement of the false corners between frames would give an incorrect result.

It is worthwhile to reduce as many false positives in the detected features as possible. A method to do this is proposed below, and evaluation results are presented in Sect. 4.1.

False corners typically arise when edges from different FOVs intersect in the OM image. Thus, they are typically located at the crossings of edges, whereas real corners are typically located at the junctions of edges. In a binary image of edges - obtained by summing the x and y gradients of the OM image - false corners will tend to be wherever there are line crossings, while real corners will tend to be wherever multiple lines, typically 2 or 3, terminate at a point. Convolving the binary image with a 7x7 ring-shaped kernel (with 1's at the perimeter of the kernel), will result in false corners having values of at least 4. These false corners are then masked from the feature detector so that they are not detected in the image.

3.4. 2-FOV Monocular

We tried two approaches for 2-FOV OM monocular SLAM: using epipolar constraints and spectral filtering.

3.4.1 Approach #1: Epipolar Constraint

The relationship between common feature points of two consecutive frames can be related using homography or essential matrix models. These models can be estimated from 7 or 8 common features respectively. This can be coupled with a random sampling consensus (RANSAC) method which tries to minimize the cumulative distances of all features from their estimated epipolar lines.

Our approach uses a similar method for trying to remove the FOV ambiguity of our OM camera. We calculate the essential matrix from 8 common features between two consecutive frames and assume that they are all from a single FOV in each frame (FOV A to FOV A, FOV B to FOV B, FOV A to FOV B, or FOV B to FOV A). We check all possible combinations of epipolar lines using our knowledge of each FOV's rotational offset and intrinsic matrix. We then find the FOV combination for each feature that minimizes the cumulative distance from their respective epipolar lines. This process is done via RANSAC for hundreds of iterations. Figure 6 shows how we would have incorporated this approach into ORB-SLAM2. This approach was not fruitful, but we believe with further refinement it may become successful. Results of the FOV estimations are shown in section 4.2.

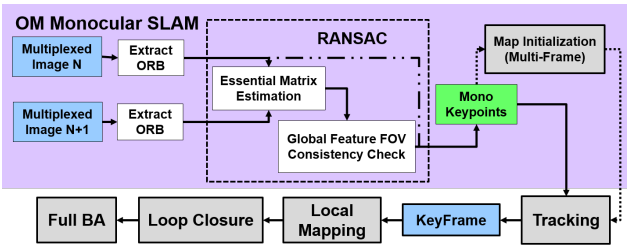

Figure 6. Overview of 2-FOV OM Extension

3.4.2 Approach #2: Spectral Filters (Anaglyph)

With spectrally filtered FOVs, disambiguation of the FOVs can be achieved by selecting color channels that correspond to their respective color filters. Since we are using cyan and red filters on our imagery, the first FOV can be extracted from the green and blue channels, and the second FOV can be extracted from the red channel. Once we extract the FOVs we can run two instances of the ORB-SLAM2 tracker to generate trajectories and point clouds. Figure 7 shows our integration into the ORB-SLAM2 architecture. Results are shown in section 4.2.

We attempted to integrate both FOVs into a single ORB-SLAM2 tracker, but we ran into difficulties with the Local Mapping/Tracker. This was due to the way the graph algorithm adds KeyFrames which can cause confusion for the graph search algorithm when two distinct, non-overlapping FOVs are added sequentially.

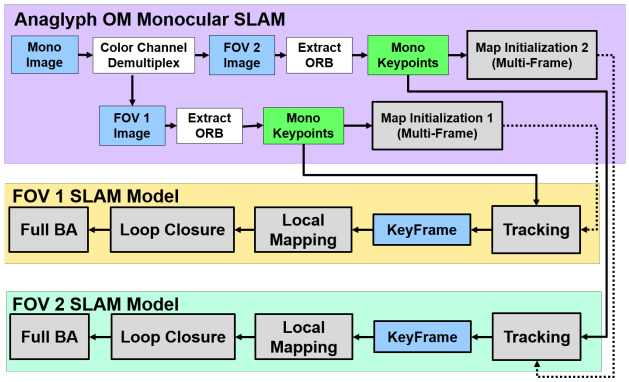

Figure 7. Overview of 2-FOV Anaglyph Extension

3.5. Stereo

We tried the same two approaches for OM stereo SLAM: using epipolar constraints and spectral filtering.

3.5.1 Approach #1: Epipolar Constraint

Our first approach extended the stereo ORB-SLAM2 implementation to infer FOVs using epipolar constraints. Analogous to the traditional stereo case, epipolar constraints can be used in OM imagery to find correspondences. The problem is largely the same, as the FOVs remain offset like in stereo vision but are now projected onto the same focal plane.

The geometry of the cameras constrains correspondences between OM FOVs such that pairs of correspondences must lie along an epipolar line. Given a normalized image coordinate p_L, the corresponding coordinate in the other FOV, p_R, must satisfy

$$p_L E^T p_R = 0, \tag{1}$$

where E is the essential matrix of the 2-FOV system. By modifying the correspondence search within ORB-SLAM2 to search within the same image, we were able to identify stereo correspondences in OM images (Figure 8).

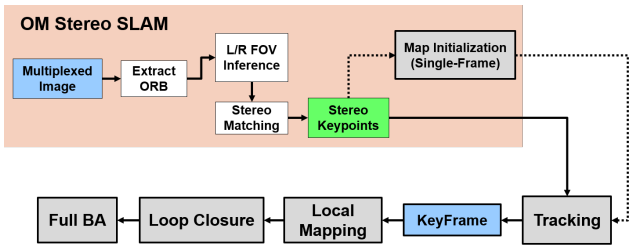

Figure 8. Overview of OM Stereo extension. Note the L/R FOV inference step in addition to normal feature extraction, which incorporates epipolar constraint matching.

3.5.2 Approach #2: Spectral Filters

Similar to the spectrally filtered monocular OM case, spectral filtering improves disambiguation of stereo FOVs by encoding additional information into the images. Using the same demultiplexing approach as described in Section 3.4.2, the two FOVs can be recovered and passed into ORB-SLAM2's standard stereo tracking, with the blue/green color channels as the left image and red channels as the right image.

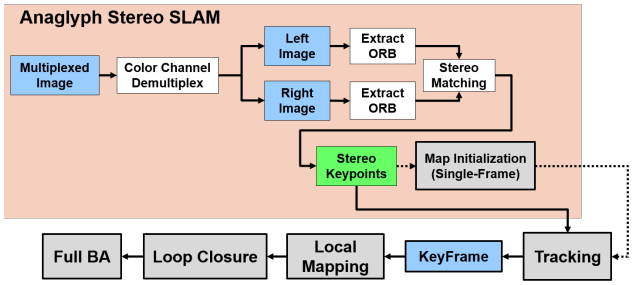

Figure 9. Overview of OM anaglyph stereo extension. FOV estimation in this approach relies on color channel demultiplexing, as opposed to epipolar constraints.

4. Experimental Results

4.1. Feature Extractor Performance

The false corner filtering technique described in Section 3.3 is successful in reducing the number of false corner detections in OM images for the ORB feature detector, but it does not have a significant effect for other feature detectors, as shown in Figure 10 (right). We hypothesize that it may be due to the ORB detector being more sensitive to these false corners. Additionally, it has the downside of removing many valid feature detections from the OM image (Figure 10, left), particularly those that are located on objects with high spatial frequency and many edges, such as a painting or a vase with reflections on its surface. This may be due to the way false corners are classified in the binary image of edges. The criterion that false corners have a convolution greater than 3 with the ring kernel will give false positives in those high-frequency regions of the image. Those features are valid and should not be masked from the feature detector. Further work is needed to refine the method of distinguishing between false and valid features.

This false corner filtering technique serves as a proof of concept and was not implemented in the ORB-SLAM2 code base.

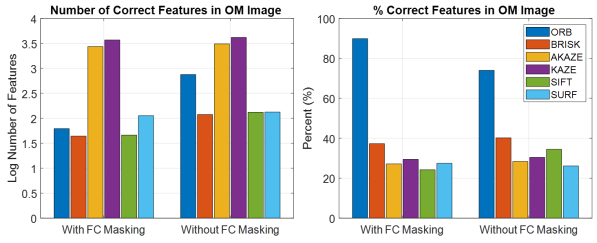

Figure 10. Feature extractor performance with/without false corner masking. Correct features are those that match the ones detected in the individual FOVs.

4.2. 2-FOV Monocular

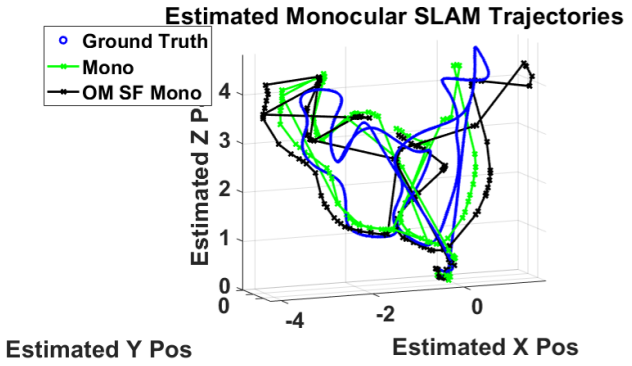

Figure 11. 2-FOV trajectory estimations. The SF monocular system performed better than the baseline (non-multiplexed) monocular system.

Our first 2-FOV monocular approach, which infers correspondences' FOVs between OM frames using epipolar constraints, needs further refinement. Without any additional information from spectral filtering, RANSAC is unable to consistently estimate the correct FOV for these correspondences (Figure 12).

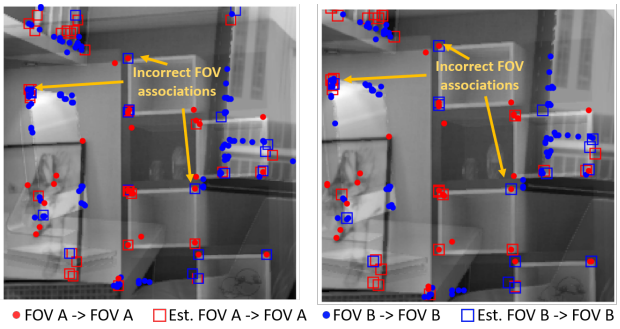

Figure 12. Feature FOV estimations from consecutive frames using our OM monocular with epipolar constraints approach. As shown in the image, there are too many incorrect FOV associations to perform SLAM.

Our second 2-FOV monocular approach implemented spectral filtering to provide more information to disambiguate the FOVs. The estimated trajectory generated using this approach had a RMSE of 0.1750 meters as compared to the ground truth trajectory, which is in family with traditional single FOV monocular performance (Table 4.3). The significance of this result is our ability to simultaneously generate two point clouds on the same trajectory, which enables the OM system to map the environment at a faster rate and, with development could utilize more reference points when localizing the camera. Figure 13 shows the added benefits of this approach.

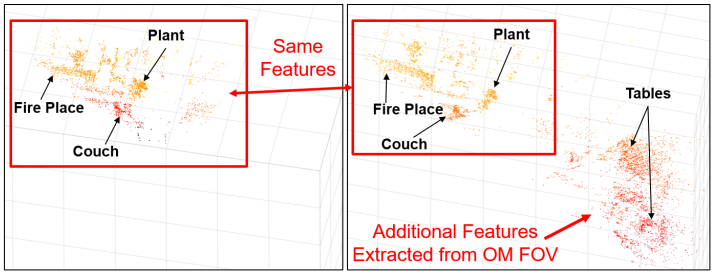

Figure 13. 2-FOV point cloud performance. The OM system is able to generate two separate point clouds with the same camera resources as the unmultiplexed system.

Table 1. RMSE (m) of scaled estimated trajectories

Mono	Stereo	Stereo OM	Stereo OM SF	Mono OM SF
0.1915	0.1217	0.1896	0.1545	0.1750

4.3. Stereo

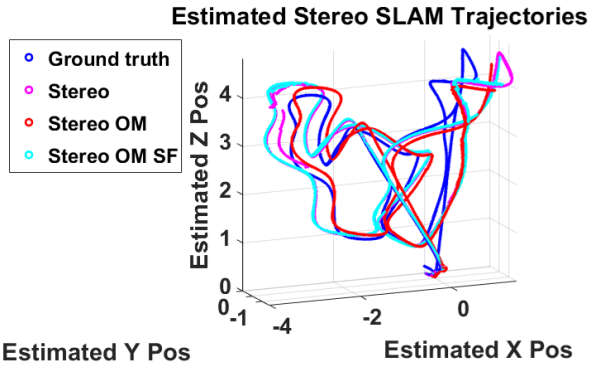

Figure 14. Stereo trajectory estimations. The unfiltered stereo OM system performed on par with the unmultiplexed case and the filtered stereo OM system had the best performance.

The 2-FOV OM stereo approach with epipolar constraints performs comparably with the standard stereo case in estimating trajectory. The RMSE of the OM system is 0.1896 meters compared to the baseline RMSE of 0.1217 meters. While the RMSE increased, this decrease in performance is offset by the savings associated with OM systems in reducing SWaP-C. This result is significant and validates the application of optical multiplexing to the field of SLAM. With spectral filtering applied to the incoming FOVs, the RMSE of the trajectory estimate improves to 0.1545 meters. By encoding additional FOV information into the images, our implementation of ORB-SLAM2 is able to better utilize the disambiguated FOVs and increase performance closer to standard stereo levels.

5. Conclusion

In this project, we sought to explore the application of OM cameras to SLAM algorithms. We developed a process to generate synthetic OM imagery, created a false corner mitigation technique for feature extraction in OM images, and implemented extensions of an open source SLAM architecture.

We were successful with enhancing ORB-SLAM2 to work with spectral filtered synthetic imagery for both stereo and 2-FOV monocular OM cameras. We were also successful with performing SLAM with a gray scale stereo OM camera by using known geometric constraints. These cameras were in family with traditional non-multiplex camera performance.

Achieving consistent success in disambiguation of the FOVs for a 2-FOV OM monocular camera was not fruitful. The additional challenge associated with this configuration requires further refinement of our approach.

Our reference code can be found at https://github.mit.edu/ri25017/6.869_Project.

6. Personal Contribution

My first contribution was the preparation and modifications of the initial ORB-SLAM2 and FlightGoggles codebases so our team had the tools needed for our OM modifications. I also integrated the OM monocular color filter extension into ORB-SLAM2, as well as the implementation of the OM monocular epipolar constraint approach in Matlab. We all collaborated on the approaches for data synthesis methods/models.

References

[1] Andover corporation: Dichroic filter spectral data. https://www.andovercorp.com/products/dichroic-filters/dichroic-filter-spectral-data. Accessed: 2018-12-12.

[2] Flea3 usb3 vision cameras. https://www.ptgrey.com/flea3-usb3-vision-cameras. Accessed: 2018-12-12.

[3] Vinay shah. private communication.

[4] L. C. S. K. V. S. A. Suleiman, Z. Zhang. Navion: A 2mw fully integrated real-time visual-inertial odometry accelerator for autonomous navigation of nano drones. *IEEE Journal of Solid State Circuits (JSSC), VLSI Symposia Special Issue*, 54(4), 2019.

[5] A. Antonini, W. Guerra, V. Murali, T. Sayre-McCord, and S. Karaman. The blackbird dataset: A large-scale dataset for uav perception in aggressive flight. In *2018 International Symposium on Experimental Robotics (ISER)*, 2018.

[6] R. J. W. D. Floreano. Science, technology and the future of small autonomous drones. *Nature*, 521(7553):460–466, 2015.

[7] V. S. Y. R. R. H. S. Emma Landsiedel, Corrie Smeaton. Refractive optically multiplexed lwir imaging system. *Proc. SPIE, Advanced Optics for Defense Applications: UV through LWIR III*, 10627:22432–22445, 2018.

[8] R. H. S. et al. Design architectures for optically multiplexed imaging. *Opt. Express*, 18:22432–22445, 2010.

[9] M. K. K. M. N. O. P.-A. P. S. S. H. T. R. J. Wood, B. Finio and J. P. Whitney. Progress on 'pico' air vehicles. *The International Journal of Robotics Research*, 31(11):1292–1302, 2012.

[10] T. Sayre-McCord and W. e. a. Guerra. Visual-inertial navigation algorithm development using photorealistic camera simulation in the loop. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[11] R. H. S. Y. Rachlin, V. Shah and T. Shih. Dynamic optically multiplexed imaging. *Proc. SPIE*, 9600(96003), 2015.