Sensor Applications

Robust Optical Spatial Localization using a Single Image Sensor

Ivan White*, Deva K. Borah***, and Wei Tang**

Klipsch School of Electrical and Computer Engineering, New Mexico State University, Las Cruces, NM, 88003, USA

** Student Member, IEEE*

*** Member, IEEE*

**** Senior Member, IEEE*

Manuscript received XXX XX, 2019; revised XXX XX, 2019; accepted XXX XX, 2019. Date of publication XXX XX, 2019; date of current version XXX XX, 2019.

Abstract—A novel three-dimensional (3D) spatial localization method using a single camera with temporal-difference image processing is proposed. The proposed active localization method uses a ring of light-emitting diodes (LEDs) embedded on the target. The diameter and the central location of the ring's image on the image sensor are used to estimate the target location using a Volterra series expansion of the target coordinates. No knowledge of the camera hardware parameters is needed. Instead, Volterra series parameters are obtained through a prior training that needs to be performed only once. The proposed method can be implemented with low computational complexity and storage. The performance of the proposed method is compared against an earlier method that relies on prior knowledge of the camera hardware parameters. The proposed method demonstrates excellent performance even when the target is located far away from the axial direction of the camera lens. The proposed method can be applied in low-power outdoor unmanned aerial vehicle localization and indoor robotic navigation.

Index Terms—Image Sensor, Object Localization, Single Image 3D Localization, unmanned aerial vehicle

I. INTRODUCTION

Optical spatial localization has advantages in low-power applications compared to other wireless localization methods such as ultrasound and radio-frequency (RF) methods [1]–[3]. The main techniques in optical localization include the received signal strength (RSS) [4], the time-of-flight (TOF) [5] and the angle-of-arrival (AOA) methods [6], [7]. An optical localization technique can be either active or passive. In the active method, the optical source of illumination is placed on the target object or on the anchors (beacons). The optical sensor, i.e., a camera, takes pictures of the optical source and performs image processing in order to obtain the spatial location of the target object. The active method is limited by the variation of the optical detector performance, the background light intensity, and the length of the baseline in the stereo-camera systems [6], [7]. On the other hand, a passive method does not use an active optical source. It uses the optical parallax principle [8] and triangulation techniques to obtain spatial information. This approach requires high computational overhead to process the reflections from the target. The passive method is also difficult to be applied in a 3-D application. In general, the performance of localization techniques is measured in terms of accuracy, range, as well as the implementation complexity and computational overhead. A comprehensive review of optical localization methods for low power sensor networks is given in [9].

In flying drone applications, although the global positioning system (GPS) or ranging sensors are useful, there have been interests in using camera-based localization methods [10]. The detection range in the drone applications should reach tens of meters with an accuracy in the order of centimeters [11]. The camera-based methods include drone-camera and ground-camera methods. In the drone-camera method [12]–[14], the drone takes a video using its on-board camera and transmits the video to the base station. The video contains landmark information so that the base station is able to compute the location of the drone using the landmark information. This typically requires extensive image processing algorithms, e.g., extended Kalman Filters. The drone-camera methods are restricted to a pre-fixed arena with landmarks. They suffer from high computational overhead at the base station as well as high power consumption at the drone due to video transmission, thus reducing the battery life of the drone. In the ground-camera method, the camera takes pictures of the drone from the ground or from a base station while the drone is equipped with special optical markers. For example, [15] installs infrared LEDs on the drone so that a ground robot can take a video. The distance between the drone and the ground robot is 5-meter. In another example, [16] applies both thermal and visible image processing to localize the drone. The infrared or thermal imager in the ground-camera method has a lower resolution compared to a visible light optical imager, which may limit the detection range and accuracy.

In [17], an improved ground-camera 3-D spatial localization method using a single temporal difference image sensor is proposed. The method uses a light-emitting diode (LED) ring that can be embedded in a drone. The advantage of using the LED ring as a marker is that the diameter of the ring can always be obtained from the major axis of the ring's image, which is an ellipse, regardless of the relative orientation of the target. The approach [17], referred to as the lens geometry method (LGM) in this paper, obtains the spatial localization using the image shape based on the lens geometry. The lens equations are highly accurate when the target is located close to the axial direction of the lens. Therefore, LGM's performance deteriorates when the

Corresponding author: Wei Tang (e-mail: wtang@nmsu.edu).

Associate Editor:

Digital Object Identifier

Fig. 1: Example UAV application scenario. Left: LEDs in the ring are off; Middle: LEDs in the ring are on; Right: temporal difference of the two images (color inverted).

target is located much away from the axial direction. To address this problem, in this paper, we propose a novel method based on a Volterra series expansion of the target location. The proposed Volterra series method (VSM) provides excellent performance even when the target is located much away from the axial direction. Further, VSM does not require knowledge of the camera parameters and relies on a prior training that needs to be performed only once. We present a detailed comparison between LGM and VSM. The rest of the paper is organized as follows. Section II presents the experimental system. In Section III, we describe the proposed VSM algorithm. Results and discussions are given in Section IV. Finally, Section V provides the conclusions of the paper.

II. EXPERIMENTAL SYSTEM

In order to compare the performance between VSM and LGM methods, the experimental setup used is the same as described in [17] for the LGM method. The LED ring with a diameter of 2 feet (61cm) blinks at a frequency of 20 Hz to represent a medium-sized unmanned aerial vehicle (UAV). To capture the image of the LED ring, a camera is held by a tripod adapter, which is placed on a rotating swivel stand. The rotation angles are marked by a protractor so that the camera is able to change its horizontal view angle of the target LED ring. The vertical view angle is adjusted by turning the camera vertically using the tripod adapter. The camera resolution is 2976 by 2976 pixels. The true distance between the camera and the LED ring is measured using a digital laser meter.

The camera takes pictures of the LED ring and computes the temporal difference image by calculating the difference of the value of each pixel between the current picture and the previous picture. Thus, if A and A' denote the pixel values at two consecutive sampling times for the same pixel, then the temporal difference image is obtained by finding $|A - A'|$ for all pixels. This removes static interfering signals. Multiple consecutive temporal difference images are used to distinguish the image of the ring from other interfering signals. Toward this end, a pixel is determined as a possible part of the LED ring only when it is observed to be "blinking", i.e., when the amplitude difference for that pixel in the temporal difference images is higher than a pre-defined amplitude threshold. Next, the pixel's value must change with a predetermined frequency to be acceptable as part of the ring's image. This can be implemented in hardware by adding pixel values of multiple temporal difference images and accepting only those pixels with a sum value within two specified threshold values. Since the background is not changing according to the target frequency, only the target LED is distinguished from the image. This processing removes random flickering noise from the environment or other movements of objects, including the drone

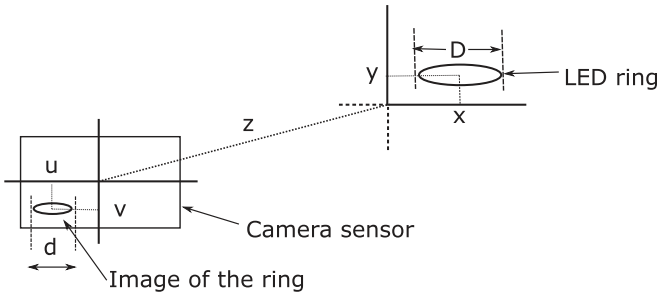

Fig. 2: Geometry of object localization. The xy and uv planes are parallel to each other with no rotation. The z axis is normal to both xy and uv planes, and it intersects the uv plane at the origin $(0, 0, 0)$.

itself. Only the pixels representing the LED ring are needed to run the VSM algorithm to be described next. An application scenario of using temporal image difference is shown in Fig. 1.

III. PROPOSED VSM ALGORITHM

The geometry of the optical localization system is illustrated in Fig. 2. From the image, the values of u, v and d in terms of the number of pixels are measured and used in the VSM algorithm as described below. Let (x_i, y_i, z_i) be the location of the LED ring center with respect to the center of the camera lens during the i-th measurement. Three quantities are measured from the image: 1) the ring diameter (d_i) in the image plane, 2) the shift of the image ring center from the camera center along the x-direction (u_i), and 3) the shift of the image ring center from the camera center along the y-direction (v_i). In the following, we consider only the magnitude of these quantities, since the sign can be easily interpreted from the direction of the image shift from the camera center. In an ideal camera, x_i and y_i are linearly related to u_i and v_i respectively for a given distance z_i, and z_i is inversely dependent on d_i as $z_i = fD/d_i$, where D is the true LED ring diameter and f is the focal length of the camera lens. In practice, the camera lens has distortions, and the inter-dependence of x_i, y_i and z_i on u_i, v_i and d_i is more complex. To handle this realistic scenario, we propose to use a nonlinear model given by a Volterra series. In general, such a series can be accurately represented using N terms, where N is reasonably large. In our case, to keep the complexity low, we use a second order series with $N = 10$ terms. The proposed series representation is given by $x_i = \alpha_x^{(0)} + \alpha_x^{(1)} u_i + \alpha_x^{(2)} v_i + \alpha_x^{(3)} (1/d_i) + \alpha_x^{(4)} u_i^2 + \alpha_x^{(5)} v_i^2 + \alpha_x^{(6)} (1/d_i^2) + \alpha_x^{(7)} u_i v_i + \alpha_x^{(8)} u_i/d_i + \alpha_x^{(9)} v_i/d_i$, where $\alpha_x^{(0)} = 1$ and $\alpha_x^{(i)}$, $i = 1, \cdots, 9$, are parameters to be determined. Note that x_i not only contains linearly dependent terms on u_i, v_i and $1/d_i$ but also the non-linear terms, e.g., u_i^2, and the cross-terms, e.g., u_i/d_i. We can similarly present expressions for y_i and z_i in terms of coefficients $\alpha_y^{(i)}$ and $\alpha_z^{(i)}$, $i = 0, \cdots, 9$, respectively. Therefore, a total of 27 parameters are to be determined.

Parameter determination: For a given camera, the parameters, α_x, α_y, and α_z are fixed, and thus they need to be calculated only once. To determine the parameters, a set of L measurements is obtained by placing the LED ring at multiple locations. For each measurement, the LED ring is placed at a known specific location (x_i, y_i, z_i) and the values of u_i, v_i and d_i are obtained from the corresponding image. Let us define a vector of parameters for x_i as $\mathbf{p}_x = [\alpha_x^{(0)} \; \alpha_x^{(1)} \; \cdots \; \alpha_x^{(9)}]^T$, and a vector of measured quantities as $\mathbf{s}_i = [1 \; u_i \; v_i \; 1/d_i \; u_i^2 \; v_i^2 \; 1/d_i^2 \; u_i v_i \; u_i/d_i \; v_i/d_i]^T$, so that the

corresponding Volterra series expression is $x_i = \mathbf{s}_i^T \mathbf{p}_x$, where $(\cdot)^T$ denotes a transpose operation. We can similarly define $\mathbf{p}_y$ and $\mathbf{p}_z$ in terms of $\alpha_y^{(j)}$ and $\alpha_z^{(j)}$, $j = 0, \cdots, 9$, respectively, and can write $y_i = \mathbf{s}_i^T \mathbf{p}_y$, and $z_i = \mathbf{s}_i^T \mathbf{p}_z$. From the set of L measurements, we construct the measurement matrix $\mathbf{S} = [\mathbf{s}_1 \quad \mathbf{s}_2 \quad \cdots, \mathbf{s}_L]^T$, so that $\mathbf{s}_1^T$, $\mathbf{s}_2^T, \cdots, \mathbf{s}_L^T$ form the L rows of the matrix $\mathbf{S}$. The corresponding location vectors are defined as $\mathbf{x} = [x_1 \quad x_2 \quad \cdots x_L]^T$, $\mathbf{y} = [y_1 \quad y_2 \quad \cdots y_L]^T$, and $\mathbf{z} = [z_1 \quad z_2 \quad \cdots z_L]^T$, that allow us to write $\mathbf{S}\mathbf{p}_x = \mathbf{x}$, $\mathbf{S}\mathbf{p}_y = \mathbf{y}$, and $\mathbf{S}\mathbf{p}_z = \mathbf{z}$. Note that the matrix $\mathbf{S}$ and the vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are known from the L training measurements. Therefore, the parameter vectors $\mathbf{p}_x$, $\mathbf{p}_y$ and $\mathbf{p}_z$ are determined using $\mathbf{p}_x = \mathbf{S}^\dagger \mathbf{x}$, $\mathbf{p}_y = \mathbf{S}^\dagger \mathbf{y}$, and $\mathbf{p}_z = \mathbf{S}^\dagger \mathbf{z}$, where $\mathbf{S}^\dagger = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T$ is the matrix pseudo-inverse. This requires $L > 10$ independent training measurements. Note that parameter determination *does not* involve real-time operations. The 27 real parameters are obtained *a priori*, and stored for the given camera system.

Unknown location estimation after parameters are estimated by training. When the ring is positioned at an unknown location (x_k, y_k, z_k), we measure u_k, v_k and d_k from its image, and construct the measurement vector $\mathbf{s}_k = [1 \quad u_k \quad v_k \quad 1/d_k \quad u_k^2 \quad v_k^2 \quad 1/d_k^2 \quad u_k v_k \quad u_k/d_k \quad v_k/d_k]^T$. The unknown location is then estimated as $\hat{x}_k = \mathbf{s}_k^T \mathbf{p}_x$, $\hat{y}_k = \mathbf{s}_k^T \mathbf{p}_y$, and $\hat{z}_k = \mathbf{s}_k^T \mathbf{p}_z$, where the hat denotes an estimate. This requires at most 27 multiplications and 27 additions on a real time basis. Note that some of the parameters may be zero causing a reduction in the computational complexity.

Using built-in pixel value subtracting circuits, the VSM algorithm can be implemented in a hardware-friendly way with the help of the temporal difference image sensor [18]. The temporal difference operation can also be performed in the computing domain when using a regular image sensor. The temporal difference image sensor can also adjust the difference threshold to adapt to different background noise and illumination contrast of the target circle. Finally, as outlined in [17], the registration of the target requires finding the minimum area rectangle that encloses the circle or ellipse in the image, which requires much less computation power compared to a registration of a full ellipse or special shape that may rotate, potentially requiring computationally intensive Hough Transform. This saves the processing energy for battery-powered devices.

IV. RESULTS AND DISCUSSION

We obtain a set of 60 measurements by placing the LED ring at various locations in an indoor environment at distances of 1m, 3m, 10m, and 30m. This distance range has importance for UAV applications, for example, in docking. We divide the available data into two groups: (1) a set of $L = 40$ training data samples ensuring that enough training exposure to various locations is obtained. The training data set is used to estimate the parameter vectors $\mathbf{p}_x$, $\mathbf{p}_y$ and $\mathbf{p}_z$ as described in the previous section. (2) a set of 20 test data samples for testing the performance of the proposed algorithm. Our experimental results are analyzed in Figs. 3, 4 and 5 as described below.

In Fig. 3, we display the 3D localization performance of the proposed VSM algorithm and compare its performance against LGM. Observe that LGM performs quite accurately when the image of the target is near the center of the image sensor. However, for large x and y, i.e., when the image is located at the edge of the image sensor, LGM's performance degrades rapidly (e.g., points A and B in Fig. 3) as the lens equations used in LGM become inaccurate due

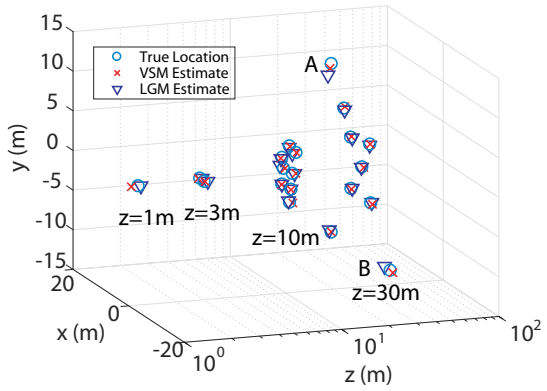

Fig. 3: Experimental results of the true and estimated 3-D locations of the target.

to imperfections of the optical system. On the other hand, due to the more general model used in VSM, its performance remains superior at the edges.

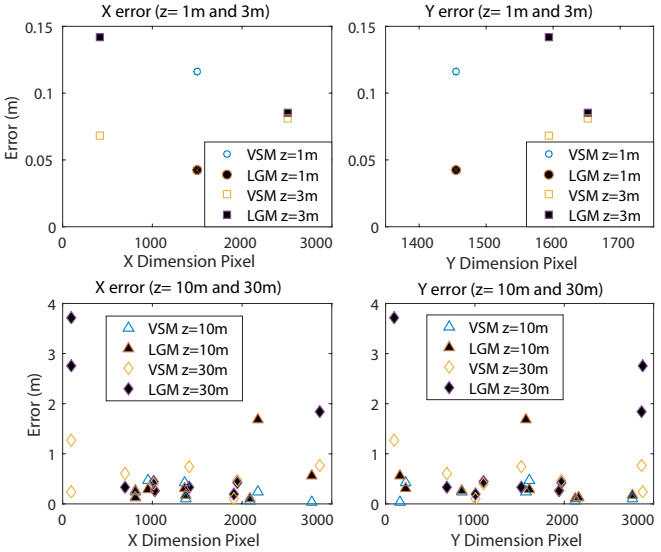

Fig. 4: Error between true and estimated results when the centroid of the target is at different locations on the images.

We demonstrate the estimation error on x and y for various z values for both LGM and VSM in Fig. 4. The estimation error for the k-th location is calculated as $\epsilon_k = \sqrt{(x_k - \hat{x}_k)^2 + (y_k - \hat{y}_k)^2 + (z_k - \hat{z}_k)^2}$, where (x_k, y_k, z_k) is the true location of the LED ring center and $(\hat{x}_k, \hat{y}_k, \hat{z}_k)$ is the estimate obtained using LGM or VSM. Note that VSM's error stays reasonable for all cases, while LGM's performance significantly worsens toward the edge of the image sensor. Defining the overall mean-squared error (MSE) as $(1/K)\sum_{k=1}^{K} \epsilon_k^2$, we have obtained an MSE of 1.45 m^2 for LGM and 0.21 m^2 for VSM. Since LGM performs accurately when the LED ring center is near the camera axis, a hybrid method (HM) can be employed, wherein LGM is first employed to estimate the LED ring location, and if the $x-$ or $y-$ axis value is found to be more than 1.5 m, then HM switches to VSM. We have noticed that HM improves the performance over both LGM and VSM at the cost of increased computational complexity, and the MSE of HM for our case is found to be 0.182 m^2. In general, for large z, the error performance degrades for all the methods as the image gets smaller.

In Fig. 5, we analyze the average estimation errors on x, y and z

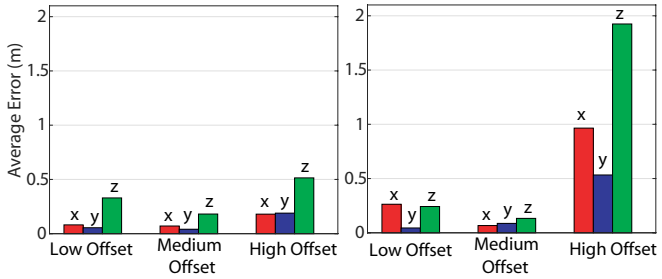

Fig. 5: Error behavior for low, medium and high offsets of the target.

co-ordinates for small, medium and large axial offsets of the target. We define axial offset as $r = \sqrt{x^2 + y^2}$, which represents the radial distance of the target with respect to the camera axial direction. The target is considered to have a low axial offset if r is less than 20% of z, i.e., $r < 0.2z$. Similarly, a medium offset arises when $0.2z \leq r < 0.4z$, and a large axial offset refers to the case when $r \geq 0.4z$. The average of the estimation errors, $|x - \hat{x}|$, $|y - \hat{y}|$ and $|z - \hat{z}|$ for the x, y and z axes respectively, show that at low axial offsets, both VSM and LGM perform well for x, y and z. In fact, the estimation of z is found to be better for LGM. However, for large offsets, VSM outperforms LGM for all the axes.

We compare VSM against LGM in Table I. The two methods have their own benefits and drawbacks. Although the real-time operations for VSM is high, the complexity can be *significantly reduced* by setting all the small parameters to zero. In our experimental results, it is found that setting all the parameters, $\alpha_x^{(i)}$, $i = 1, \cdots, 9$, to zero except the most dominant one, reduces the number of operations to 5 multiplications, 1 division, and 3 additions with a storage requirement of only 3 real numbers. The MSE degrades slightly to 0.45 m^2, which is still far better than the LGM's MSE of 1.45 m^2.

Table 1: Comparison of LGM and VSM. LGM stores tan(·) function values for angles $\leq 50^o$ with 5 divisions per degree.

Attribute	LGM [17]	VSM (This Work)
Prior training	Not needed	Needed
Accurate values of camera parameters	Needed	Not needed
Real-time computations	7 multiplications, 3 divisions, 5 additions & 1 square root	31 multiplications, 4 divisions & 27 additions
Storage	250 real numbers	27 real numbers
Error performance close to sensor center	Very good	Performance is sometimes slightly worse than LGM
Error performance at sensor edges	Performance degrades	Performance is much better than LGM

V. CONCLUSION

A novel 3-D spatial localization method based on the Volterra series is presented in a system that uses a blinking LED ring marker and a digital camera receiver. Unlike many other optical triangulation AOA methods that require multiple receivers, the proposed method uses the temporal image difference from only a single imager. The method can accurately localize a target over a wide range of the camera's field of view and has potential applications in outdoor UAV flight control, for example, docking of UAVs, and in indoor robotic navigation. The method is required to store a set of 27 parameters, which can be reduced to just 3 parameters for low complexity efficient hardware implementation. The stored parameters are used along with the measured image quantities to estimate the target location. The proposed method does not require any prior knowledge of the camera hardware parameters. On the other hand, if the camera hardware parameters are known, the proposed method can be combined with LGM to further improve the localization performance.

ACKNOWLEDGMENT

This work is sponsored by United States National Science Foundation grant 1652944.

REFERENCES

[1] Y. Liu, X. Yu, S. Chen, and W. Tang, "Object Localization and Size Measurement Using Networked Address Event Representation Imagers," *IEEE Sensors Journal*, vol. 16, no. 9, pp. 2894–2895, May 2016.

[2] S. Sivathasan and D. C. O'Brien, "RF/FSO Wireless Sensor Networks: A Performance Study," in *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, Nov 2008, pp. 1–5.

[3] X. Jin and J. F. Holzman, "Differential Retro-Detection for Remote Sensing Applications," *IEEE Sensors Journal*, vol. 10, no. 12, pp. 1875–1883, Dec 2010.

[4] S. Yang, H. Kim, Y. Son, and S. Han, "Three-Dimensional Visible Light Indoor Localization Using AOA and RSS With Multiple Optical Receivers," *Journal of Lightwave Technology*, vol. 32, no. 14, pp. 2480–2485, July 2014.

[5] O. Sgrott, D. Mosconi, M. Perenzoni, G. Pedretti, L. Gonzo, and D. Stoppa, "A 134-Pixel CMOS Sensor for Combined Time-of-Flight and Optical Triangulation 3-D Imaging," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 7, pp. 1354–1364, July 2010.

[6] M. S. Islam and R. Klukas, "Indoor Positioning Through Integration of Optical Angles of Arrival with an Inertial Measurement Unit," in *Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium*, April 2012, pp. 408–413.

[7] S. Linga, B. Roy, H. Asada, and D. Rus, "An Optical External Localization System and Applications to Indoor Tracking," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2008, pp. 1127–1132.

[8] "Laser Range Finder 28044: Laser Diode and CMOS Camera (Detector)," Parallax Inc, CA, USA, Tech. Rep., 2012.

[9] N. Iliev and I. Paprotny, "Review and Comparison of Spatial Localization Methods for Low-Power Wireless Sensor Networks," *IEEE Sensors Journal*, vol. 15, no. 10, pp. 5971–5987, Oct 2015.

[10] M. Achtelik, A. Bachrach, R. He, S. Prentice, and N. Roy, "Stereo vision and laser odometry for autonomous helicopters in GPS-denied indoor environments," *Proceedings of SPIE*, vol. 7332, 733219, 2009.

[11] J. Kang, K. Park, and H. Kim, "Analysis of Localization for Drone-Fleet," in *2015 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct 2015, pp. 533–538.

[12] L. Jayatilleke and N. Zhang, "Landmark-based Localization for Unmanned Aerial Vehicles," in *2013 IEEE International Systems Conference (SysCon)*, April 2013, pp. 448–451.

[13] S. Zhao, Z. Hu, M. Yin, K. Z. Y. Ang, P. Liu, F. Wang, X. Dong, F. Lin, B. M. Chen, and T. H. Lee, "A Robust Real-Time Vision System for Autonomous Cargo Transfer by an Unmanned Helicopter," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 2, pp. 1210–1219, Feb 2015.

[14] J. Skoda and R. Bartak, "Camera-Based Localization and Stabilization of a Flying Drone," in *Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference*, 2015, pp. 372–377.

[15] M. Faessler, E. Mueggler, K. Schwabe, and D. Scaramuzza, "A Monocular Pose Estimation System Based on Infrared LEDs," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 907–913.

[16] Y. Tang, Y. Hu, J. Cui, F. Liao, M. Lao, F. Lin, and R. S. H. Teo, "Vision-Aided Multi-UAV Autonomous Flocking in GPS-Denied Environment," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 1, pp. 616–626, Jan 2019.

[17] I. White, E. Curry, D. K. Borah, S. J. Stochaj, and W. Tang, "An Optical Spatial Localization Algorithm Using Single Temporal Difference Image Sensor," *IEEE Sensors Letters*, pp. 1–4, 2019.

[18] S. Chen, W. Tang, X. Zhang, and E. Culurciello, "A 64x64 Pixels UWB Wireless Temporal-Difference Digital Image Sensor," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 12, pp. 2232–2240, Dec 2012.