A Spatial Localization and Attitude Estimation System for Unmanned Aerial Vehicles Using a Single Dynamic Vision Sensor

Hunter Stuckey, *Student Member, IEEE*, Amer Al-Radaideh, Liang Sun, *Member, IEEE*, and Wei Tang, *Member, IEEE*

***Abstract*—This paper presents a three-dimensional (3D) localization and attitude estimation system to track a Unmanned Aerial Vehicle (UAV) using a single camera without prior knowledge of the environment. The hardware system consists of a Dynamic Vision Sensing (DVS) camera, a circle-shaped blinking marker made by Light-Emitting Diodes (LEDs), and a base station computer. The algorithm for spatial localization and attitude estimation includes a temporal video filter, triangulation-based location and attitude estimation, and 3D real-time plotting with a graphical user interface (GUI). The temporal video filter processes the image stream from the DVS camera to identify the frequency of the marker and removes the background image. The circle-shaped marker creates an ellipse in the image, whose diameter length and angles are utilized for calculating the location and attitude of the UAV, which offers a low-computing overhead. The proposed system has been evaluated in hardware flight testing. The results are compared with the benchmark data from both the infrared motion capturing system for localization and the on-board inertial measurement units of the UAV for attitude estimation. The accuracy and detection range surpasses similar state-of-the-art systems. The proposed method provides a simple yet accurate solution for tracking the location and attitude of a UAV.**

***Index Terms*—Localization, attitude estimation, unmanned aerial vehicle, dynamic vision sensor.**

I. Introduction

UNMANNED Aerial Vehicles (UAVs) have been widely witnessed in applications that are either expensive or dangerous for humans [1]. Examples of these applications include but are not limited to aerial photography, building safety inspection, precision crop monitoring, weather surveillance, and delivery [2]–[5]. In such applications, attitude estimation and spatial localization of the UAVs are two critical tasks. Here attitude estimation refers to obtaining the three-dimensional (3D) orientation (i.e., roll, pitch, and yaw angles) of a UAV with respect to the local navigation coordinate system of the UAV itself, while spatial localization is defined as the task of acquiring the 3D location information relative to a reference point, at which an observer or the base station is located. These measurements are critical for UAV coordination and control, especially in critical flight phases, such as take-off and landing [6]. The system performing such tasks is expected to be operating in real-time while maintaining reasonable accuracy. Moreover, it should also be both reliable under various scenarios and easy to operate by users or control interfaces. Furthermore, since most mobile operations are supported by batteries, low power consumption is also an important requirement for quality system performance.

Conventional attitude estimation is usually achieved by the inertial navigation system (INS) using on-board gyroscopes and accelerometers [7], [8]. While INS provides a low complexity solution, the integration of data from these sensors is prone to drift due to error accumulation without external corrections. In order to obtain long-term accuracy and stability, attitude estimation is also obtained using the combined solution of INS and the Global Positioning System (GPS). Although the fused data from both INS and GPS can provide reliable and precise results for pose estimation, GPS signals are not always available and its accuracy degrades

Manuscript received 25 May 2022; revised 23 June 2022; accepted 27 June 2022. Date of publication 7 July 2022; date of current version 1 August 2022. This work was supported in part by the United States National Science Foundation under Grant ECCS-1652944, in part by the New Mexico Space Grant Consortium (NMSGC), and in part by the National Aeronautics and Space Administration (NASA) Cooperative under Agreement 80NSSC20M0034. The associate editor coordinating the review of this article and approving it for publication was Dr. Brajesh Kumar Kaushik. *(Corresponding author: Wei Tang.)*

Hunter Stuckey and Wei Tang are with the Klipsch School of Electrical and Computer Engineering, New Mexico State University, Las Cruces, NM 88003 USA (e-mail: hss127@nmsu.edu; wtang@nmsu.edu).

Amer Al-Radaideh and Liang Sun are with the Department of Mechanical and Aerospace Engineering, New Mexico State University, Las Cruces, NM 88003 USA (e-mail: radaideh@nmsu.edu; lsun@nmsu.edu).

Digital Object Identifier 10.1109/JSEN.2022.3187423

in environments such as battle fields and urban canyons [9]. Moreover, the above methods are only for a UAV to obtain its own attitude. If an observer or operator needs to estimate the attitude of a UAV, it would rely on the wireless communication link between the UAV and the observer, which is also not only always reliable but also energy-consuming [10]. Therefore, a direct attitude estimation method is preferred for an observer to obtain the pose of a UAV without wireless communication.

Spatial localization of a UAV is also typically performed using either onboard INS sensors or off-board sensing technologies, referred to as self-localization and cooperative localization [11]–[14]. Typical onboard sensors include GPS, RF transceivers, and cameras, while off-board sensors usually rely on the infrastructure pre-installed in the working environment, such as landmarks or motion capture systems. The methods for spatial localization can also be categorized into active and passive ones according to signal types and communication directions [15]. In an active method, an UAV generates signals, and an observer (base station) analyzes and tracks the signal using certain sensing and processing technologies. For example, the images or videos taken from a UAV can be analyzed to find its location with the help from known landmarks. In a passive method, an observer generates signals, such as images, videos, sonar signals, and radio signals, to track the UAV location [16]–[18]. In applications where a direct wireless data link between a UAV and its base station is not available and reliable, it is expected to apply a passive method to obtain the spatial location of the UAV.

Image-based localization and attitude estimation methods have been attracting researchers due to the recent advancement of image processing technologies. While an on-board image sensor consumes significant power for image sensing, processing, or transmitting, off-board vision-based localization and attitude estimation methods have gained popularity in recent years [19], [20]. This is because off-board sensing and processing do not rely on the limited onboard power of a UAV. Typical off-board vision systems in such applications use multilocular sensors, i.e., sensors with more than one camera, which are fixed at known locations. For example, a stereo vision system uses two cameras [21] and certain localization methods would need three cameras [22]. Monocular vision-based methods have also been proposed, which use only one camera and the known marker constellation on the UAV [23]–[25]. The single-camera methods have a wider range of applications such as using the follower UAV to track the leader UAV. In this paper, we focus on UAV spatial localization and attitude estimation using a novel monocular vision-based method.

The main design considerations of the proposed monocular vision-based UAV localization and attitude estimation systems include detection range, detection accuracy, system power consumption, system computing overhead, reliability, and computing time [26], [27]. In typical design specifications [24], [25], [28], [29], the accuracy of localization is on the level of centimeters while the accuracy of attitude estimation is usually about 5 degrees. In order to precisely analyze the location and pose of the UAV using images and videos, the sensor system should be able to operate within 30 meters. Besides the above specifications, the hardware and software methods should achieve real-time estimation. For example, when a UAV is flying in slow motion at the level of around 1 meter per second, the system latency should be less than 0.1 seconds to make sure the location accuracy is less than 10 cm. This makes the challenge for the computing hardware and processing algorithm since the input video data volume is high. A simple algorithm is expected with lower computing overhead to save the processing time as well as the computing power.

In the application of optical localization and pose estimation, one of the primary challenges is to locate the key component of a UAV from the background image [30]. A typical method is adding several Light-Emitting Diodes (LEDs) as the markers on the UAV. The LED markers have unique colors and form a constellation so that the image processing algorithm can find them in the image. This paper applies a monocular vision-based optical method for spatial localization and attitude estimation using a circle-shaped blinking LED ring as the marker on the UAV. The LED ring is blinking at a high frequency to distinguish itself from the background. We use an event-based dynamic vision sensor (DVS) to capture only the blinking marker from the image. The circle-shaped marker helps in developing a low computing overhead trigonometric algorithm for localization and attitude estimation. Finally, the system estimates the pose and the 3D location of the UAV relative to the camera.

This paper extends from our prior work [23] of the DVS-based optical localization system, which was validated using still testing and short-range testing for localization. The main unique contributions of this paper include (1) a novel attitude estimation algorithm, (2) long-range localization testing, and (3) improved optical image processing filters. In specific, the attitude estimation algorithm uses the length ratio between the major and minor axes of the ellipse in the image to calculate the roll and pitch angles, which is one of the first algorithms proposed for monocular attitude estimation for circle-shaped LED marker, to the best of our knowledge. Also, the long-range testing of the localization method validates that the proposed system can achieve reliable localization results in a 25-meter range, which was not validated in our prior publications. In addition, the optical filter algorithm has been improved by using erosion-dilation filters to reduce random noise from the background, which makes the system more reliable in the experimental environment. In summary, this paper presents a comprehensive method and measurement results for the proposed system that overcomes the problems in our prior work.

The remaining of the paper is organized as follows. Section II introduces the related work of this paper, including similar systems in the references and the sub-modules for the system, as well as our prior publications. Section III describes the related hardware implementation including the sensor, the marker, and the computing device. Section IV presents the software algorithm development including image pre-processing methods and the algorithms for localization and pose estimation. Section VI presents the experimental results for both long-range localization and pose estimation. Section VII discusses the advantages and shortcomings of this

work as well as future work directions. Section VIII concludes this paper.

II. Related Work

The related work of the proposed system includes dynamic vision sensors and monocular image processing algorithms for UAV spatial localization and attitude estimation. A typical optical method in such applications uses prior knowledge of the markers attached to the UAV to obtain the attitude and location. The required prior knowledge includes the shape, size, and color of the marker. The LEDs used on the marker usually form a constellation with a special shape. For instance, [25] presents an ultraviolet UV marker that forms a hexagonal constellation for a UAV leader-follower tracking application. Infrared markers are applied in [24] with a pentagonal constellation. Nevertheless, the constellation of the marker introduces ambiguity when solving the attitude of the UAV, and it also increases the computing complexity since lots of trigonometric functions are involved. In addition, the special color (UV/IR) chosen in these applications may be interfered by environmental noise since it is possible that the background also contains a similar UV/IR source. Therefore, in our work, we focus on improving the system performance using new shapes and features of the markers for optical localization and pose estimation.

Monocular vision-based UAV tracking systems have gained popularity compared to multi-camera systems [19]. For instance, a monocular vision sensor is combined with Radio Frequency Identification (RFID) and to track 3D location using 2D images with triangulation-based method in [31]. [32] focuses on robot pose estimation using a high frame rate camera and graphic processing units (GPU) running a monocular vision algorithm. The primary challenge in these methods is that they rely on image processing algorithms with a high computing overhead since the input image data have large sizes. One possible solution is to shift some computing tasks into the image sensor, such as in DVSs. DVSs have been widely used in high-speed real-time object tracking applications [33], [34]. A DVS directly records the location of the pixel that has its value changed above a certain threshold without performing a rolling shutter like in a regular camera [35], [36]. This provides a unique feature of obtaining the temporal difference image while saving comparing time and sensing energy. With this feature, only moving objects are captured in the image while the background image is removed, which provides an effective means for UAV tracking with both spatial localization and attitude estimation. In this paper, we apply a monocular DVS for real-time UAV spatial localization and attitude estimation.

In our prior work [23], we have demonstrated the solution of using a circle-shaped LED ring to perform spatial localization of a UAV. Temporal difference image processing was applied to identify the blinking LED circle from the background image. Since the diameter of the circle marker, which has a known length, can always be seen as the major axis of the ellipse in the image, the diameter of the circle can be applied as a reference length to translate the physical distance into the number of pixels in the image [28]. We also proposed correction algorithms to improve the robustness of the proposed method [29]. However, since the processing speed of a regular image sensor is too slow, no dynamic flight testing was performed in [28], [29]. The results of [24], [25] were obtained from drone hovering using a webcam and offline processing. For real-time dynamic flight tracking, there are significant challenges to implementing the proposed localization algorithm on hardware. These challenges come from the image sensor processing speed, the processing time of the localization algorithm, and the time for plotting the trace. Although [23] achieved low-speed flight testing, the UAV was flying within a 5-m distance to the camera. No long-distance testing was made and no algorithm for attitude estimation was presented.

Compared to other similar works [24], [25], our proposed method avoids potential noise interference from other infrared sources in the background image. This is because the proposed system applies a temporal image filter to match the blinking rate of the LED marker. Since the temporal filter is targeting at a specific blinking frequency, in a natural environment, it would be a much lower probability to have a noise source with the same blinking rate of the LED marker, compared to the chance of having a noise source with a similar spectrum of the color-based marker such as UV/IR/RGB-based markers. Another advantage of the proposed system is that the algorithm for localizing the circle-shaped marker demands a lower computing overhead than a marker constellation. Therefore, in this work, we demonstrate a DVS-based monocular vision system using blinking LED markers and temporal image filters to achieve simultaneous UAV localization and attitude estimation. The detailed methods that achieve reasonable accuracy with low computing overhead are presented in the following sections.

III. System Implementation

The sensing system consists of the circle-shaped blinking LED marker attached to the UAV, the DVS camera, and a base station computer connected to the DVS camera. The frequency of the blinking LED marker is controlled by a microcontroller on the UAV. The DVS camera tracks the target that blinks at the frequency of the marker using the temporal image filters implemented in the base station computer. By locating the marker LED ring, the spatial location and the attitude estimation are then performed using trigonometric algorithms in real-time.

A. Dynamic Vision Sensor

In this system, we applied a Celex4 dynamic vision sensor developed by Hillhouse Technology [37] as the main image sensor, illustrated in Fig. 1 (A). The DVS generates a stream of pixel events using the address-event representation format. A pixel event is recorded only when the same pixel has a difference in the values of two consecutive readings and the difference value is higher than a pre-defined threshold. This is performed using a pixel-level circuit that always subtracts the pixel values from their prior values and compares the difference with the threshold. The DVS can achieve an event rate of up to 500 frames/second that provides the map of pixels with significant changes between every consecutive two

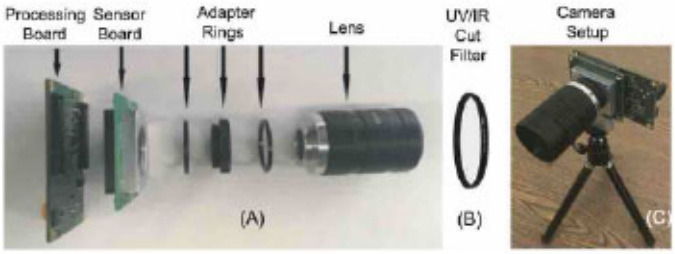

Fig. 1. (A) Picture of the Celex4 event camera, retrieved from celepixel sdk manual [37], (B) Picture of 37 millimeter UV/IR cut filter placed in front of the event-camera, (c) Camera Setup.

Fig. 2. Graphical User Interface of the Celex Dynamic Vision Sensor showing both normal image and temporal difference image of the blinking LED Marker.

Fig. 3. Picture of UAV with blinking LED marker ring.

frames. This feature is important for saving the computing overhead for the following image processing algorithm. In our implementation, an optical UV/IR cut filter is applied to remove infrared noise from the background, as shown in Fig. 1 (B). The DVS camera is connected to a Spartan-6 FPGA board XEM6301 developed by Opal Kelly Inc. through a Universal Serial Bus (USB) 3.0 interface. The system setup is shown in Fig. 1 (C) The system can display both normal and temporal-difference images using the GUI as shown in Fig. 2. In the experimental setup, the event frame time of the DVS camera is set at 4 ms.

B. Markers and Base Station

The marker installed on the UAV is a circle-shaped LED tube with a two-feet diameter, as shown in Fig. 3. The LED marker is powered and controlled by an Arduino Nano Microcontroller Board that can provide a current of 100 mA. The pre-designed blinking frequency is set at 40 Hz. The blinking frequency and the known diameter of the LED circle provide essential data for the image processing algorithms to perform spatial localization and attitude estimation. There are several important advantages of using this blinking LED ring compared to the LED constellation using UV/IR. Firstly, the DVS can easily remove the background image and identify the LED ring using a counter-based temporal band-pass image filter. It is more robust than UV/IR because the background image is less possible to contain an object that has the same blinking frequency. Secondly, the diameter of the ring can be easily translated into the number of pixels in the image to calculate the UAV altitude and the distance between the UAV and the camera, which greatly reduces the computing overhead compared with the methods using a marker constellation. The blinking LED marker generates additional weight on the UAV, which results in additional power consumption, like in other active localization methods that use optical markers. The weight of the LED ring plus the frequency control circuit in our proposed system is only 80 grams, which is 6.6% of the total weight of the UAV (1200 grams). Note that the weight of such devices could be significantly reduced by using lighter materials and a refined design. In addition, the LED ring could be powered off when not used for localization.

The base station in the proposed system is a desktop computer running Ubuntu 16 LTS that implements the image filter algorithms as well as the localization and pose estimation algorithms. The base station also performs real-time plotting of the 3D location of the UAV with the roll and pitch angles in a GUI, which is written by C++. The base station runs two threads: one is for the image filters and the other is for calculating spatial localization and pose estimation parameters. The measured latency of the base station processing is less than 100 ms.

IV. Processing Algorithm

A. Image Filters

The image processing algorithm running on the base station detects the LED ring from the input image of the dynamic vision sensor and calculates the spatial location and attitude of the UAV. Multiple image filters are applied to perceive the blinking pattern of the LED ring while removing the background image and noise pixels, which are elaborated in the first subsection. The output of the image filters is an ellipse that represents the LED marker. The most top, bottom, left, and right points of the ellipse are applied to calculate the location and attitude of the physical marker ring using a trigonometric algorithm presented in the second subsection. The image filters and geometry algorithms are running in parallel in the system. The final output of the program is the real-time 3D tracking and attitude angles that display on a GUI. The software system is programmed in C++ running on a Linux computer. The computation was performed on a CPU since the current version of the Celex camera does not compatible with the GPU on the base station.

The image filters include the temporal band-pass filter, the corrosion-dilation filter, and the output medium filter. The input of the image filters is the event-stream generated

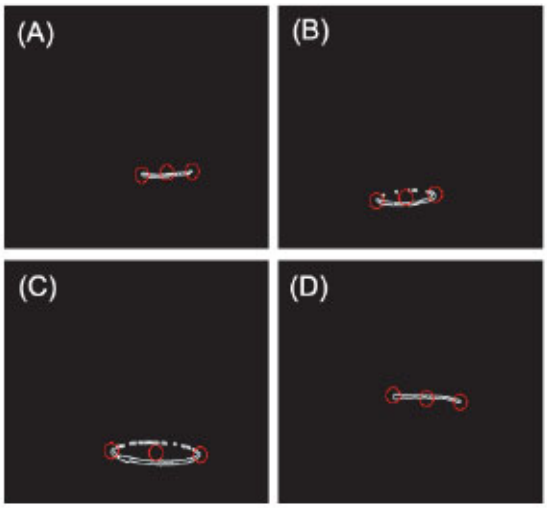

Fig. 4. Binary images obtained from temporal band-pass filter with marked left, right, and center points of target. (A) The UAV is about level with the event camera. (B,C) The UAV is below the event camera and the ring of LEDs is visible. (D) The UAV is at a slightly higher altitude above the camera.

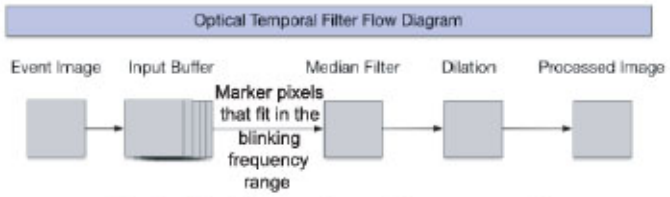

Fig. 5. Flow diagram of event-image processing.

from the DVS, which contains all the pixels whose temporal difference value is beyond the preset threshold. The temporal band-pass filter processes the event stream and identifies the pixels that belong to the blinking LED marker. This is performed using prior knowledge of the blinking rate of the LED marker. In our design, if a pixel is consistently blinking in a certain frequency range that covers the blinking frequency of the LED marker, we then consider this pixel belongs to the LED marker. Otherwise, the pixel is removed by the image filter. The output image of the temporal band-pass filter contains the image of the LED marker and some random noise pixel values. The corrosion-dilation algorithm is used to remove the random noises. After that, the key points of the marker ellipse are obtained for calculating the location and attitude of the UAV. The key points include the most top, bottom, left, and right pixels of the ellipse. The medium filter tracks the location of the key points. If some of the key points move too fast, the result is considered as the wrong detection and removed by the medium filter. This guarantees that the image processing algorithm is robust to noise pixels from the dynamic vision sensor. Example outputs of the image filters are shown in Fig. 4.

The detailed structure and operation of the image filters are shown in Fig. 5. Since the image filters are expected to identify the frequency of the blinking LED marker, the image sensor should have a higher sampling rate than the blinking frequency. To identify each on and off phase of the LED marker, the sampling rate should be twice the blinking frequency. Since the events are generated by subtraction between consecutive frames in the DVS, the total sampling rate of the DVS is at least four times the blinking frequency. Moreover, during flight testings, the moving UAV requires an even faster sampling rate for real-time tracking. The input image frames from the DVS camera are stored in a fixed-size buffer. The buffer is a First-In-First-Out memory that works as a sliding window. All the frames in the window are added together by each pixel to form a combined frame. A pixel that blinks more frequently in all the frames has a higher value in the combined frame. The pixel value in the combined frames is compared with upper and lower threshold values, which serves as the blinking frequency selector or a temporal band-pass filter. It only selects the pixels that the blinking frequency is between the two frequency thresholds. The background image and noise pixels that do not meet the blinking frequency requirement are filtered out by this temporal band-pass filter.

Algorithm 1 Temporal Band-Pass Algorithm

```
procedure TFilter(buffer,bufferFull)
    if bufferFull then                          ▷ If buffer is full, then threshold
        img = Σ_{i=0}^{L} img_buffer            ▷ images are binary
        if amount of events is within the threshold then
            set given pixel to 1
        else
            set given pixel to 0
        end if
    end if
    Post Processing on Image
    img = medianBlur(img, kernelSize)
    img = dilation(img, structElement)
    return img                                  ▷ Image Forwarded to 3D localization
end procedure
```

The temporal band-pass filter has a set of design parameters including the length of the sliding window and the frequency threshold. These parameters can be adjusted by the sampling rate of the DVS camera, the blinking frequency of the LED marker, and the expectation of the system latency. For instance, with an N frame/sec sampling rate of the DVS camera, if the length of the sliding window is W frames, and a pixel in the combined frame has a value of n, which means n events in W frames, the estimated event rate is $N * n/W$ events/sec, and the blinking rate of the pixel is $N*n/2W$ Hz. We note that the clocks between the DVS camera and the LED marker are not synchronized, the blinking measurement may vary between different sliding windows. Therefore, we apply two frequency thresholds to decide if the blinking rate is within the target range. A longer length of the sliding window results in a higher frequency resolution of the detected event rate but causes a longer latency in the system. If the shorter window length may introduce more noise pixels in the image, which may affect the accuracy of the localization and attitude estimation.

In our design, the sampling rate of the DVS camera is set at 250 frames/sec, while the blinking frequency of the LED marker is set at 40 Hz. The window length is 20 frames to assure that the total processing latency is less than 0.1 seconds.

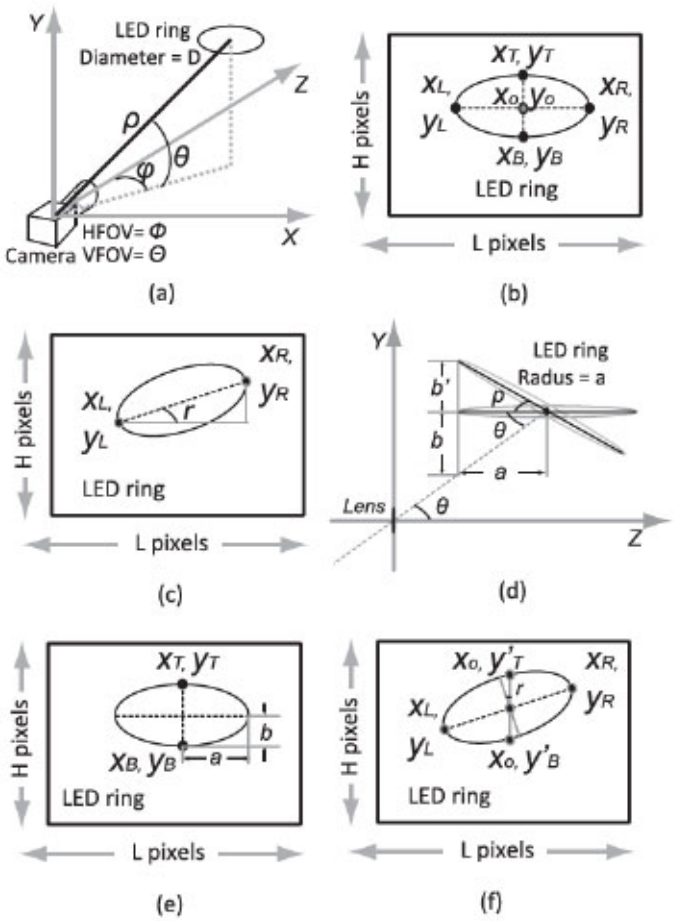

Fig. 6. Calculation of localization and attitude angles. (a) Geometry for localization and attitude estimation, (b) Measurement of the key points, (c) Estimation of the roll angle, (d) - (f) Estimation of the pitch angle.

The output of the image filters is sent to the trigonometry algorithm module for calculating spatial localization and attitude parameters. The program runs a two-threading process so it can perform image filtering and geometry calculation concurrently.

B. Spatial Localization

The geometry localization method has been proposed and implemented in [28]. As shown in Fig. 6 (a), the camera takes images of the LED marker circle located at a spherical coordinate location (radial distance ρ, inclination angleθ, azimuth angle ϕ). In a real application, key system parameters can be measured and obtained, including the physical diameter of the LED circle, D, the horizontal field of view (HFOV), Φ, and the vertical field of view (VFOV) of the camera, Θ. In the image of the LED circle as illustrated in Fig. 6 (b), assuming the target has a horizontal attitude relative to the camera, the image of the circle becomes a horizontal ellipse. The leftmost pixel and the rightmost pixel locations are (X_L, Y_L) and (X_R, Y_R). Assume that the image has L pixels in width and H pixels in height, using the known parameters (Θ, Φ, D, L, H) and the measured parameters (X_L, Y_L, X_R, Y_R) from the image, the location of the target circle (ρ, θ, ϕ) can be calculated as follows:

Firstly, the center of the target circle in the image is obtained as:

$$X_O = (X_L + X_R)/2 \tag{1}$$

$$Y_O = (Y_L + Y_R)/2 \tag{2}$$

Then the azimuth angle ϕ can be calculated using the following equations:

$$\tan\phi = \frac{x_O}{f} \tag{3}$$

where f is the focal distance of the image sensor, which may not be known exactly since we can apply another equation for the horizontal FOV Φ and the width of the image in pixel L

$$\tan\frac{\Phi}{2} = \frac{L/2}{f} \tag{4}$$

Using Equations (3) and (4), we can cancel f and have the tangent value of azimuth angle ϕ

$$\tan(\phi) = 2 \cdot X_O \cdot \frac{\tan(\Phi/2)}{L} \tag{5}$$

Similarly, the elevation angle θ can be obtained as

$$\tan(\theta) = 2 \cdot Y_O \cdot \frac{\tan\Theta/2}{H} \tag{6}$$

Finally, the radial distance, ρ, which is the distance between the centroid of the target ring and the lens of the camera is obtained as

$$\rho^2 = \left(\frac{D}{2} \cdot \frac{X_O}{|X_O - X_R|}\right)^2 \cdot \left(1 + \frac{1}{\tan^2\phi}\right) \tag{7}$$

Note that the above calculations actually do not include computing triangulation functions since $\tan\Phi$ and $\tan\Theta$ are constants and can be obtained before computing, this greatly saves computing overhead. However, in order to evaluate errors and compare them with other references, we can convert the spherical coordinates to Cartesian coordinates using the following equations:

$$\begin{cases} x &= \rho * \sin\phi * \cos\theta \\ y &= \rho * \sin\theta \\ z &= \rho * \cos\phi * \cos\theta \end{cases} \tag{8}$$

C. Attitude Estimation

The attitude of the UAV is calculated using the image of the circle-shaped marker implemented on the UAV. Since the DVS obtains only the image of the blinking marker, the circle-shaped marker projects an oblique ellipse in the image as shown in Fig. 6 (a). When the UAV is flying or hovering horizontally, the major axis of the ellipse is parallel to the horizontal axis (x-axis) in the image. The pixel location of the top (X_T, Y_T), bottom (X_B, Y_B), left (X_L, Y_L), and right (X_R, Y_R) of the ellipse in the image can be easily obtained as shown in Fig. 6 (b). These points are used to estimate the attitude of the marker. The roll angle, r, can be calculated as the ratio between the horizontal distance and vertical distance of the two terminals on the major axis of the ellipse, as shown in Fig. 6 (c) as:

$$\tan(r) = \frac{Y_R - Y_L}{X_R - X_L} \tag{9}$$

We note that with a roll angle, the most left and right pixels of the ellipse in the image are not exactly the terminals of the major axis of the ellipse. However, when the roll angles are not large, the most left and right pixels can still be applied for approximation in order to reduce the processing power.

The pitch angle is measured using the ratio between the major-axis radius and minor-axis radius of the ellipse in the image. First, we consider a simple scenario that the UAV is located on top of the camera and both roll and pitch angles are zero. In such a case, the top pixel (X_T, Y_T), and the bottom pixel (X_B, Y_B) can be directly obtained from the image to measure the radius along the y-axis b while the radius along the x-axis a is obtained from the most left pixel (X_L, Y_L) and most right pixel (X_R, Y_R) as shown in Fig. 6 (d). Since the marker has a circle shape, the ratio between b and a is related to the inclination angle θ when the plane of the marker is parallel to the ground. Based on Fig. 6 (d) and (e), we have:

$$\tan(\theta) = \frac{b}{a} = \frac{Y_T - Y_B}{X_R - X_L} \tag{10}$$

since the inclination angle can also be obtained using the center location X_O, the vertical angle of view Θ, and the total number of pixels in the height H [28].

$$\tan(\theta) = 2 \cdot Y_O \cdot \frac{\tan(\Theta/2)}{H} \tag{11}$$

Therefore, we can obtain the radius b when the pitch angle is zero

$$2 \cdot b = \tan(\theta) \cdot (X_R - X_L) \tag{12}$$

When the pitch angle is not zero, the radius b of the ellipse has an additional value of b' as shown in Fig. 6 (d). b' is related to the pitch angle p. We have

$$2 \cdot b' = Y_T - Y_B - 2 \cdot b \tag{13}$$

from the above equations, we can obtain the pitch angle as:

$$\tan(p) = \frac{Y_T - Y_B}{X_R - X_L} - 2 \cdot Y_O \cdot \frac{\tan(\Theta/2)}{H} \tag{14}$$

The above pitch angle calculation is only valid when the roll angle of the UAV is zero. When the UAV has both non-zero pitch and roll angles, the axes of the ellipse in the image are not parallel with either the x-axis or y-axis, as shown in Fig. 6 (f). In such a case, the roll angle is calculated using the same method presented in the prior paragraph. The calculation of the pitch angle is modified since the short radius b can not be directly obtained from the top and bottom pixels of the ellipse. In such a case, we use the most left and right pixels to obtain the center location of the ellipse in the x-axis (X_O), then we search the top and bottom pixels on the ellipse that has the x-axis value of X_O, which are noted as (X_O, Y'_T) and (X_O, Y'_B). Then the short radius of the ellipse is obtained with a small-angle approximation

$$Y_T - Y_B = \cos(r) \cdot (Y'_T - Y'_B) = (1 - \frac{\tan^2(r)}{2}) \cdot (Y'_T - Y'_B) \tag{15}$$

Then we can obtain the pitch angle using Equations (14) and (15).

While the temporal image filter algorithm and the triangulation-based algorithm are able to calculate the spatial location and attitude of the UAV, several non-ideal effects should be noted to avoid glitches in the system. Firstly, although the background image has a low chance to have an object that blinks at a similar frequency to the LED marker, it is still possible that several noise pixels could appear, which could be from the reflection of the LED light. Secondly, since the UAV is moving, if the moving speed is too high, the DVS camera may not be able to detect the blinking marker since the pixels of the marker in the video stream changed its location in the image. Thirdly, the finite resolution of the image limits the detection accuracy, especially when the UAV is far away from the camera. In that case, the diameter of the marker in the image occupies fewer pixels, which may affect the calculation of the triangulation functions. Also, the distortion of the object near the edge of the image due to optical imperfectness may also cause larger errors when the UAV is near the edge of the image. To address the aforementioned issues, the Kalman filter could be introduced in future work so that the noise pixels from the background would not affect the calculation [38]. The speed limit can be improved by using a DVS camera with a higher sampling rate, such as Celex 5 [39]. A higher resolution camera may alleviate the finite resolution problem to reduce errors and extend the detection range. The distortion correction method in [29] can be applied to correct edge distortions of the camera.

Like similar recent work in UAV attitude detection [23], the above attitude estimation algorithm only addresses the roll and pitch calculation. The estimation of the yaw angles was not addressed simply because the circle-shaped LED can not be applied to find the "nose" of the UAV. In order to estimate the yaw angle, the marker can be modified so that the "nose" LED of the UAV could be distinct in the LED ring. There are several methods that can be used to achieve this. For example, the "nose" LED may be applied in a different color or blinking at a different frequency. Another method would be adding a special shape, for example, a radium stick pointing to the "nose" pixel. Each of these methods would need the modification of the signal processing algorithm.

V. Error Analysis

The localization and attitude estimation errors come from various sources in the system. The first error source is the finite resolution of the camera. Due to the finite resolution, the edge pixel location of the LED ellipse in the image is not able to precisely map the true location of the ellipse edge of an ideal image, which creates errors for both localization and attitude estimation. The relative errors increase when the UAV is far away from the camera, which causes small total number of pixels of the LED ellipse. The second error source comes from the finite sampling speed of the camera and the moving speed of the UAV. The finite sampling speed may introduce errors when the UAV is moving too fast. This error also increases when the UAV is far away from the camera while the relative size of the LED ring is small in the image. In addition, the errors from system parameters such as the error in D, $HFOV$, and $VFOV$ may also contribute to errors. These errors plus

other environmental noise create a complicated problem for estimating the errors of the system.

Theoretical error analysis can be performed by focusing on the measurement errors due to the finite resolution of the camera. Since the measured pixel location is not able to precisely map the actual physical location of the LED ring, we can model such an error as a "single-pixel" error since the error is within a one-pixel range in the image. In the worst case, the mean squared error (MSE) of pixel readings is set as "1" for X_L, X_R, Y_L, and Y_R. The MSE of the center pixel location is then estimated as

$$\sigma^2(X_O) = \frac{1}{2}\left[\sigma^2(X_L) + \sigma^2(X_R)\right] = 1 \tag{16}$$

$$\sigma^2(Y_O) = \frac{1}{2}\left[\sigma^2(Y_L) + \sigma^2(Y_R)\right] = 1 \tag{17}$$

where $\sigma^2(X_L)$, $\sigma^2(X_R)$, $\sigma^2(Y_L)$, and $\sigma^2(Y_R)$ are the MSE of the pixel locations of X_L, X_R, Y_L, and Y_R, respectively, which are all estimated as "1" pixel. The MSEs of the $\tan(\theta)$ and $\tan(\phi)$ are then estimated as

$$\sigma(\tan\phi) = 2 \cdot \frac{\tan(\Phi/2)}{L} \cdot \sigma(X_O) = 2 \cdot \frac{\tan(\Phi/2)}{L} \tag{18}$$

$$\sigma(\tan\theta) = 2 \cdot \frac{\tan(\Theta/2)}{H} \cdot \sigma(Y_O) = 2 \cdot \frac{\tan(\Theta/2)}{H} \tag{19}$$

Therefore, larger resolution (L and H) can reduce the error of the measurements in the azimuth and elevation angles. From (7), the MSE of the radial distance is estimated as

$$\sigma^2(\rho) = \frac{D^2}{4}\left[\left(\frac{1}{X_O}\right)^2 + \left(\frac{1}{X_L - X_R}\right)^2 + \frac{\sigma^2(\sin\theta)}{\sin^2\theta}\right] \tag{20}$$

When the UAV is far away from the camera, the diameter of the LED ring in the image becomes small. A smaller $|X_L - X_R|$ results in a larger relative error. From (9), the MSE of the roll angle can be estimated as

$$\sigma^2(\tan(r)) = \left(\frac{1}{Y_R - Y_L}\right)^2 + \left(\frac{1}{X_R - X_L}\right)^2 \tag{21}$$

From (14), the MSE of the pitch angle can be estimated by

$$\sigma^2(\tan(p)) = \sigma^2(\tan(r)) + \sigma^2(\tan(\theta)) \tag{22}$$

While a theoretical error analysis could be performed, we have performed a detailed error estimation and correction methods and published the results in [29]. In this work, we focus on the error from the experimental results, which shows that the relative error is acceptable for localizing the target application while estimating the attitude without using Wi-Fi or a GPS network. The primary solutions to alleviate the above errors include increasing the resolution of the camera and using high-speed cameras. Moreover, the coordinate system is referred to the location of the camera as the origin, which is moveable. There could be drift during the tracking process, therefore, applying the Kalman filter can improve the reliability of the tracking system. Due to the limited space of the paper, the experimental results of Kalman filters and the Cramer-Rao bound of the estimators will be presented in our future works.

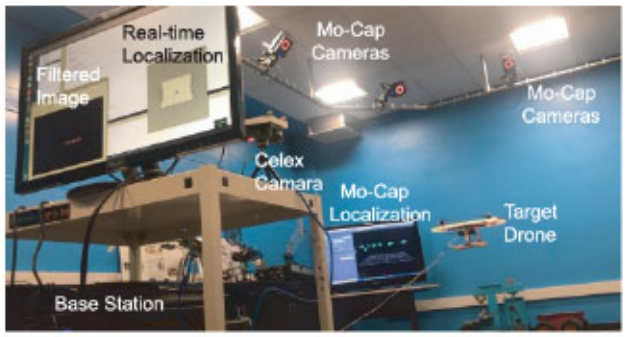

Fig. 7. Experimental environment for real-time 3D localization testing, where a flying quadcopter was being tracked by both a motion-capture system and a single dynamic vision sensor running the proposed temporal 3D localization algorithm.

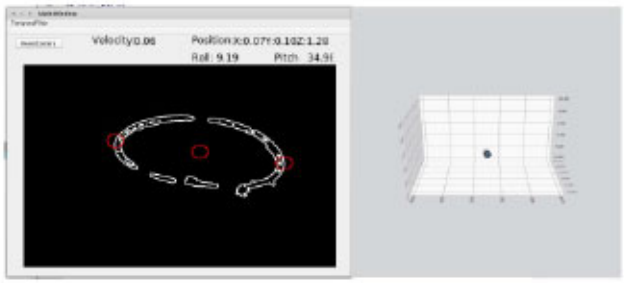

Fig. 8. The Graphical User Interface (GUI) developed for tracking localization and attitude estimation.

VI. Experimental Setup and Results

The proposed spatial localization and attitude estimation system is first characterized in a motion-capturing (MOCAP) lab. The results from the proposed system are compared with the localization results obtained from the MOCAP system that uses 10 infrared cameras with reflection markers, as well as the attitude data recorded by the onboard sensors of the UAV as benchmarks. The MOCAP flight testing arena has a size of 8 meters in length and 6 meters in width, as shown in Fig. 7. Our developed GUI plots the 3D location and displays the location coordinations and attitude angles in real-time as shown in Fig. 8. Since the MOCAP system uses infrared, which would interfere with the Celex DVS, an optical UV/IR band-pass filter was applied to the Celex camera. During the testing, the MOCAP system tracks the spatial location using the infrared reflectors attached to the UAV and the Celex Camera, so that the relative location can be referred to the Celex DVS camera as the origin. The Celex DVS camera tracks both the spatial location and attitude of the UAV using the blinking LED marker. The onboard inertial measurement units of the UAV are applied to track the attitude (roll and pitch) in real-time and compared with the result from the DVS-based sensors.

To evaluate the error of the proposed system, the calculated spatial localization data from the single-camera system were subtracted from the benchmark data from the MOCAP system to obtain the error of spatial localization. Similarly, the attitude data were compared between the results from the single-camera system and the recorded data from the onboard IMU as the benchmark. The errors were then divided by the benchmark data to obtain the relative error. In spatial

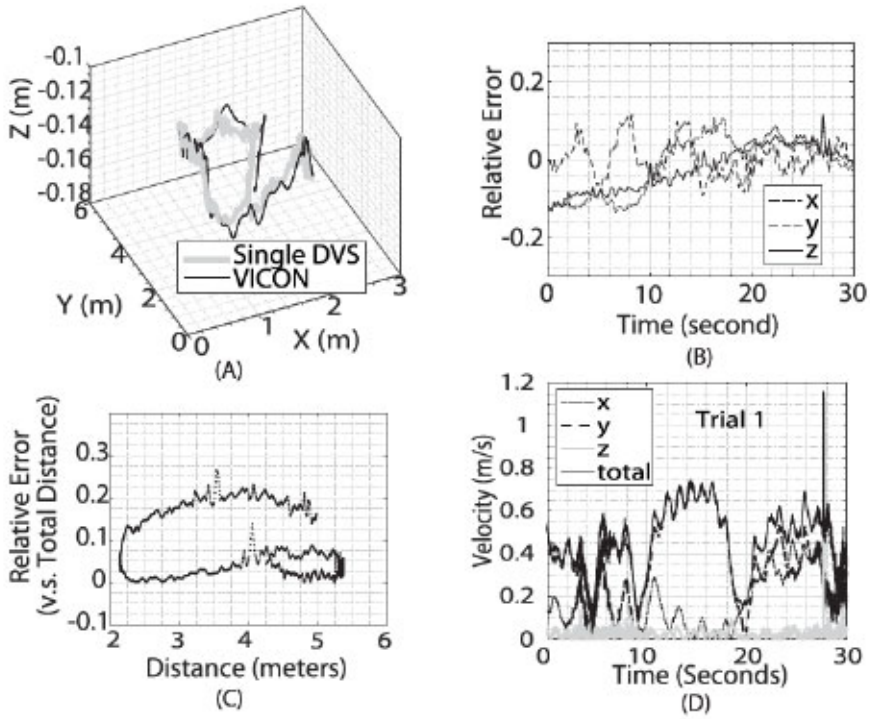

Fig. 9. Experimental results of the localization system (A) 3D traces tracked by the proposed DVS system and the benchmark VICON system (unit in meters). (B) The relative error with respect to the coordinates. (C) The relative error over distance. (D) The UAV velocity during the testing.

localization, the relative error was evaluated by the flight time, the overall distance between the UAV and the Celex camera, and the relative error for each axis. Here the x-axis refers to the horizontal direction, while the y-axis is the vertical direction. Since the Celex camera was mounted on a tripod above the ground, a negative y value means the UAV is below the Celex camera. In each axis, the average error was below 6% for the trial. The highest error is on the x-axis when the UAV is moving horizontally.

In Fig. 9 (A), the proposed single-camera temporal filtering localization 3D traces are plotted in correspondence with the Mo-Cap traces. The 3D traces from the proposed method align with the benchmark traces retrieved from the MOCAP system. The proposed system can capture the UAV trajectory and accurately identify the 3D location in real-time (100 ms latency) relative to the camera's position. Fig. 9 (B) presents the errors. The error appears to be relatively larger when the UAV was making a sharp turn or at a higher velocity. This measurement depends on both the most X-axis and Y-axis edge locations of the markers. The scatter plots in Fig. 9 (C) show the relative error over distance. The distance between the UAV and the camera (ρ) ranged between 2 and 5.5 m, which is limited by the space size that the MOCAP system covers. Fig. 9 (D) presents the 3D flight speed over time in the experiment. It has been experimentally shown that the system can effectively detect or generate data when the UAV is hovering and located at a distance of 30 m or less from the sensor [28], [29]. It is worth noting that when the UAV is in motion, the temporal difference image processing method is affected by both the blinking rate of the marker and the moving speed of the object. Therefore, the sensor is prone to miss a few frames randomly. In our experiments, a reliable

TABLE I
SYSTEM PARAMETERS AND EXPERIMENTAL PERFORMANCE OF THE PROPOSED LOCALIZATION SYSTEM

System Parameter		Experimental Performance	
Sensor Resolution	768 × 640 pixel	Localization Accuracy	<6%
Sensor Dynamic Range	90 dB	Detection Range	20 - 25 m
Sensor Frame Rate	250 frames/sec	Flight Speed	<2 m/s
Desktop CPU	AMD Ryzen 7 2700 8-core	Detection Rate	10 Hz

localization was observed when the target is moving at a speed of about 0.5 m/s, and at a distance of 20 m from the event-based sensor. This experimental result is presented in Fig. 10. The attitude measurement results compared with the benchmark data from IMUs are shown in Fig. 11.

VII. DISCUSSION

By comparing the experimental results obtained from the proposed single camera system and the benchmark results from the MOCAP system, we notice that the single-camera system performs an accurate tracking of both spatial localization and attitude of the UAV. The main advantages of the proposed system include the usage of a single camera and the low-complexity processing algorithm for simultaneous localization and attitude estimation using pixel locations of the image and the prior measured physical parameters of the LED ring and the camera system. This is achieved by the temporal image filter and the circle-shaped LED marker. Table II compares the proposed system in this paper and recently published similar systems. The proposed system achieved a longer range of detection and a lower relative error compared

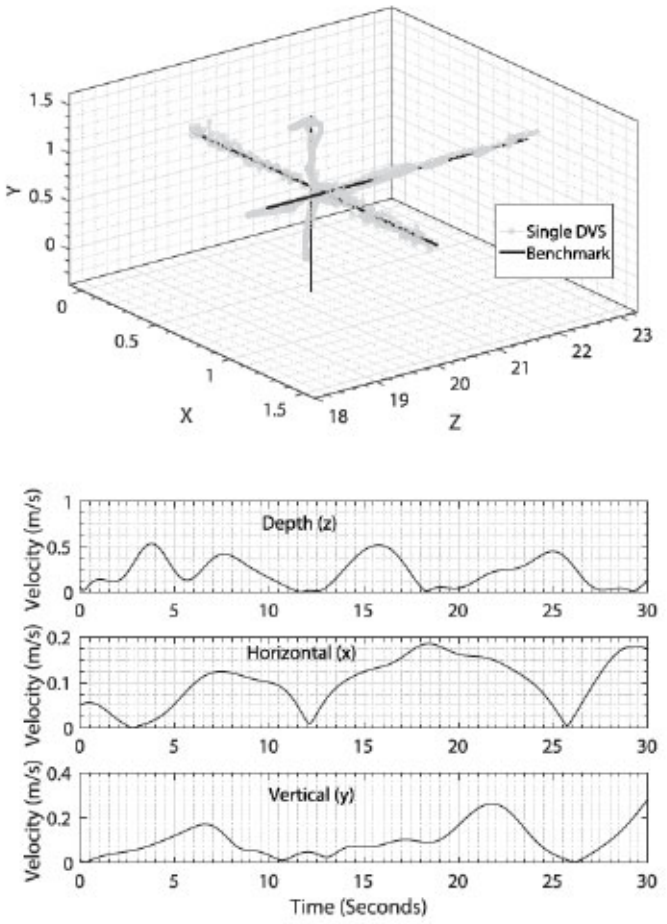

Fig. 10. Experimental results of long-distance real-time dynamic testing around 20 m (z-direction) in three directions. Left: 3D trace of the UAV and the benchmark; Right: the velocity for the testing in each direction.

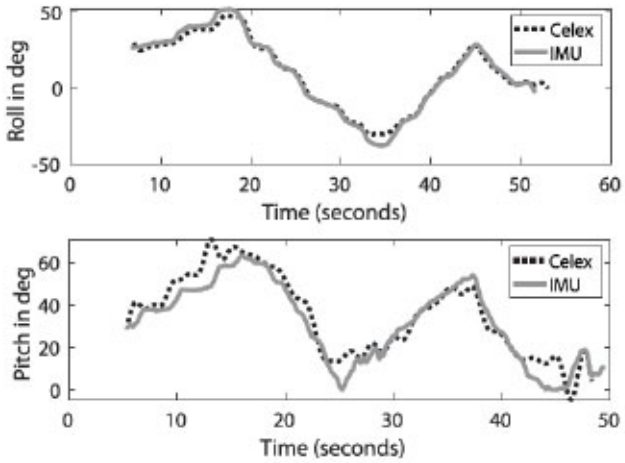

Fig. 11. Experimental results attitude measurement compared with benchmark data from IMUs.

with other methods. Since our system is using the temporal image filter to match the blinking rate of the marker instead of using a color filter, it reduces potential noise interference from the background that contains a similar color to the marker. The additional advantage of lower computing overhead is obtained by the built-in pixel-wise computing feature of the DVS imager.

The main challenges of the current system come from the limited resolution of the DVS imager and the processing time

TABLE II
COMPARISONS OF RECENTLY PUBLISHED UAV LOCALIZATION METHODS USING MONOCULAR CAMERAS AND LED MARKERS

	This Work	RA-L 2018 [24]	RA-L 2019 [25]
Sensor Type	Monocular Dynamic Vision Sensor	2xIMU + PF-MPE Camera	Monocular Ultraviolet Camera
Number of Sensors	1	3	1
Marker	Blinking LED & Circle Shape	Five RGB LEDs & Cross-Constellation	Blinking UV LEDs & Hexagon Constellation
Processing Algorithm	Temporal Filter & Optical Localization	Particle Filter & Pose Estimation PFMPE	Geometric Calculation
Detection Range	Up to 30 m	15 m	15 m
Error	2-6 cm (Range 6 m)	20 cm	15 m * 20% = 3 m
Processing Time	100 ms (10 Hz)	20 Hz	10 Hz

of the image filters. The finite resolution creates the restriction of accuracy while the processing time determines the maximum speed of the UAV that the system can track. In the future, we plan to apply a higher-speed DVS camera to solve the above problem. Moreover, by modifying the circle-shaped LED marker and the related algorithm, we are designing the yaw-angle estimation method. Furthermore, we are going to apply a Kalman filter to alleviate the motion effects for tracking high-speed UAVs. Another option for performing the localization and attitude is to perform ellipse detection algorithms [40]–[42] in the processing unit, which can directly capture the ellipse shape from the image. The computing complexity, reliability, and processing time are currently under our investigation.

VIII. CONCLUSION

In this paper, we presented a single camera method for tracking the 3D location and attitude of a multi-copter Unmanned Aerial Vehicle (UAV) based on a dynamic vision sensing camera and a circle-shaped blinking LED marker. The algorithm applies a temporal band-pass filter for image processing and a triangulation-based method for localization and attitude estimation. The signal processing system has a low computing overhead thanks to the circle-shaped marker. The flight-testing experiment results show that the proposed system has a long detection range and high accuracy compared with similar methods. Since the proposed system uses only one camera, it has the potential to be applied to low-power mobile applications for tracking UAVs without a direct wireless data link. In the future, we plan to include the estimation of the yaw angle using a modified circle-shaped marker and remove motion effects on event noise using Kalman filters.

REFERENCES

[1] L. Apvrille, Y. Roudier, and T. J. Tanzi, "Autonomous drones for disasters management: Safety and security verifications," in *Proc. 1st URSI Atlantic Radio Sci. Conf. (URSI AT-RASC)*, May 2015, pp. 1–2.

[2] M. Alwateer, S. W. Loke, and W. Rahayu, "Drone services: An investigation via prototyping and simulation," in *Proc. IEEE 4th World Forum Internet Things (WF-IoT)*, Feb. 2018, pp. 367–370.

[3] D. Joshi, "Drone technology uses and applications for commercial, industrial and military drones in 2020 and the future," *Bus. Insider*, Dec. 2019. Accessed: Jul. 3, 2022. [Online]. Available: https://www.businessinsider.com/drone-technology-uses-applications

[4] S. Banker. (Jun. 2020). *Is the Future of Drones Now?* Forbes. [Online]. Available: https://www.forbes.com/sites/stevebanker/2020/06/11/is-the-future-of-drones-now/#acf0eb032840

[5] L. Schroth. (Jun. 2020). *The Drone Market Size 2020–2025: 5 Key Takeaways*. Droneii. [Online]. Available: https://droneii.com/the-drone-market-size-2020-2025-5-key-takeaways#:~:text=1.,away,is 42.8 billion USD.&text=From generating 22.5 billion USD,almost double that in 2025.

[6] F. Liu, Z. Wei, and G. Zhang, "An off-board vision system for relative attitude measurement of aircraft," *IEEE Trans. Ind. Electron.*, vol. 69, no. 4, pp. 4225–4233, Apr. 2022.

[7] Z. Yang and S. Shen, "Monocular visual–inertial state estimation with online initialization and camera–IMU extrinsic calibration," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 1, pp. 39–51, Jan. 2017.

[8] H. Fourati, "Heterogeneous data fusion algorithm for pedestrian navigation via foot-mounted inertial measurement unit and complementary filter," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 221–229, Jan. 2015.

[9] H. Zhang, W. Yuan, Q. Shen, T. Li, and H. Chang, "A handheld inertial pedestrian navigation system with accurate step modes and device poses recognition," *IEEE Sensors J.*, vol. 15, no. 3, pp. 1421–1429, Mar. 2015.

[10] L. Gupta, R. Jain, and G. Vaszkun, "Survey of important issues in UAV communication networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1123–1152, Nov. 2016.

[11] N. Iliev and I. Paprotny, "Review and comparison of spatial localization methods for low-power wireless sensor networks," *IEEE Sensors J.*, vol. 15, no. 10, pp. 5971–5987, Oct. 2015.

[12] A. Al-Radaideh and L. Sun, "Self-localization of tethered drones without a cable force sensor in GPS-denied environments," *Drones*, vol. 5, no. 4, p. 135, Nov. 2021.

[13] A. Al-Radaideh and L. Sun, "Observability analysis and Bayesian filtering for self-localization of a tethered multicopter in GPS-denied environments," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, Jun. 2019, pp. 1041–1047.

[14] A. Al-Radaideh and L. Sun, "Self-localization of a tethered quadcopter using inertial sensors in a GPS-denied environment," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, Jun. 2017, pp. 271–277.

[15] C. Shinde, R. Lima, and K. Das, "Multi-view geometry and deep learning based drone detection and localization," in *Proc. 5th Indian Control Conf. (ICC)*, Jan. 2019, pp. 289–294.

[16] I. Guvenc, F. Koohifar, S. Singh, M. L. Sichitiu, and D. Matolak, "Detection, tracking, and interdiction for amateur drones," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 75–81, Apr. 2018.

[17] S. R. Ganti and Y. Kim, "Implementation of detection and tracking mechanism for small UAS," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, Jun. 2016, pp. 1254–1260.

[18] X. Shi, C. Yang, W. Xie, C. Liang, Z. Shi, and J. Chen, "Anti-drone system with multiple surveillance technologies: Architecture, implementation, and challenges," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 68–74, Apr. 2018.

[19] S. Srigrarom, S. M. Lee, M. Lee, F. Shaohui, and P. Ratsamee, "An integrated vision-based detection-tracking-estimation system for dynamic localization of small aerial vehicles," in *Proc. 5th Int. Conf. Control Robot. Eng. (ICCRE)*, Apr. 2020, pp. 152–158.

[20] A. Rohan, M. Rabah, and S.-H. Kim, "Convolutional neural network-based real-time object detection and tracking for parrot AR drone 2," *IEEE Access*, vol. 7, pp. 69575–69584, 2019.

[21] F. Fooladgar, S. Samavi, S. M. R. Soroushmehr, and S. Shirani, "Geometrical analysis of localization error in stereo vision systems," *IEEE Sensors J.*, vol. 13, no. 11, pp. 4236–4246, Nov. 2013.

[22] Y. Liu, X. Yu, S. Chen, and W. Tang, "Object localization and size measurement using networked address event representation imagers," *IEEE Sensors J.*, vol. 16, no. 9, pp. 2894–2895, May 2016.

[23] H. Stuckey, A. Al-Radaideh, L. Escamilla, L. Sun, L. G. Carrillo, and W. Tang, "An optical spatial localization system for tracking unmanned aerial vehicles using a single dynamic vision sensor," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 1–8.

[24] L. Teixeira, F. Maffra, M. Moos, and M. Chli, "VI-RPE: Visual-inertial relative pose estimation for aerial vehicles," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 2770–2777, Oct. 2018.

[25] V. Walter, N. Staub, A. Franchi, and M. Saska, "UVDAR system for visual relative localization with application to leader–follower formations of multirotor UAVs," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2637–2644, Jul. 2019.

[26] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source slam system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.

[27] Z. Kaleem and M. H. Rehmani, "Amateur drone monitoring: State-of-the-art architectures, key enabling technologies, and future research directions," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 150–159, Apr. 2018.

[28] 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 Lett.*, vol. 3, no. 3, pp. 1–4, Mar. 2019.

[29] I. White, D. K. Borah, and W. Tang, "Robust optical spatial localization using a single image sensor," *IEEE Sensors Lett.*, vol. 3, no. 6, pp. 1–4, Jun. 2019.

[30] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.

[31] Z. Wang, M. Xu, N. Ye, F. Xiao, R. Wang, and H. Huang, "Computer vision-assisted 3D object localization via COTS RFID devices and a monocular camera," *IEEE Trans. Mobile Comput.*, vol. 20, no. 3, pp. 893–908, Mar. 2021.

[32] Y. Liu, P. Sun, and A. Namiki, "Target tracking of moving and rotating object by high-speed monocular active vision," *IEEE Sensors J.*, vol. 20, no. 12, pp. 6727–6744, Jun. 2020.

[33] Y. Hu, J. Binas, D. Neil, S.-C. Liu, and T. Delbruck, "DDD20 end-to-end event camera driving dataset: Fusing frames and events with deep learning for improved steering prediction," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–6.

[34] S. Hoseini, G. Orchard, A. Yousefzadeh, B. Deverakonda, T. Serrano-Gotarredona, and B. Linares-Barranco, "Passive localization and detection of quadcopter UAVs by using dynamic vision sensor," in *Proc. 5th Iranian Joint Congr. Fuzzy Intell. Syst. (CFIS)*, Mar. 2017, pp. 81–85.

[35] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor," *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, Jan. 2008.

[36] C. Li *et al.*, "Design of an RGBW color VGA rolling and global shutter dynamic and active-pixel vision sensor," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2015, pp. 718–721.

[37] *Celepixel User Manual*, CelexPixel Technol., Singapore, 2019.

[38] G. Mao, S. Drake, and B. D. O. Anderson, "Design of an extended Kalman filter for UAV localization," in *Proc. Inf., Decis. Control*, Feb. 2007, pp. 224–229.

[39] J. Xu, M. Jiang, L. Yu, W. Yang, and W. Wang, "Robust motion compensation for event cameras with smooth constraint," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 604–614, 2020.

[40] T. Ellis, A. Abbood, and B. Brillault, "Ellipse detection and matching with uncertainty," *Image Vis. Comput.*, vol. 10, no. 5, pp. 271–276, Jun. 1992.

[41] K. Kanatani, "Statistical bias of conic fitting and renormalization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 3, pp. 320–326, Mar. 1994.

[42] J. Zhao, S. Huang, L. Zhao, Y. Chen, and X. Luo, "Conic feature based simultaneous localization and mapping in open environment via 2D LiDAR," *IEEE Access*, vol. 7, pp. 173703–173718, 2019.