In: IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC) 2023, 1-5 October 2023, Barcelona, Spain
DOI: 10.1109/DASC58513.2023.10311251

Enabling UAVs night-time navigation through Mutual Information-based matching of event-generated images

Naiara Escudero
Boeing Research & Technology
Cranfield University
Madrid, Spain
naiara.escudero@boeing.com

Michael W. Hardt
Airspace Operational Efficiency
Boeing Research & Technology
Madrid, Spain
michael.w.hardt@boeing.com

Gokhan Inalhan
Autonomous Systems & AI
Cranfield University
Cranfield, UK
Inalhan@cranfield.ac.uk

***Abstract*—Advanced Air Mobility is expected to revolutionize the future of general transportation. However, to make it a reality, significant challenges arise requiring technologies to ensure the expected attributes in these scenarios: resilience, robustness, large operational range, high accuracy, low SWaP equipment, and real-time processing. Although existing visual-based navigation solutions for aerial applications provide outstanding results under nominal conditions, their performance is highly constrained by the lighting conditions, making them infeasible for real operations. With the main focus of addressing this limitation, and expanding the current operational range to include extreme low-illuminated environments, this paper presents a solution which leverages one of the most powerful properties of event cameras: their high dynamic range. Thus, data provided by an event camera (also called dynamic vision sensor) is used to estimate the relative displacement of a flying vehicle during night-time conditions. To that end, two different threads running in parallel have been developed: a reference map generator, operating at low frequency, focused on reconstructing a 2-D map of the environment, and a localization thread, which matches, at high frequency, real-time event-generated images against the reference map by applying Mutual Information to estimate the aircraft's relative displacement.**

***Index Terms*—Visual-based Navigation, Event Cameras, Bio-inspired sensors, Night-time navigation, Mutual Information, Image alignment, Low-illuminated conditions.**

I. Introduction

The significant challenges associated with the concepts underlying Advanced Air Mobility (AAM) require the introduction of sophisticated solutions able to deliver the performance required for achieving safe autonomous flight operations. In this context, one of the major challenges corresponds to the demanding navigation requirements, imperative to substitute the onboard pilot. Even as technology advances, GNSS, supported by its enhancement techniques, remains an essential part of the current navigation systems. However, though GNSS-based navigation techniques provide accurate, reduced cost and easy integration solutions, they are vulnerable to radio interference as well as multi-path errors. Consequently, there is at present a huge effort focused on developing supporting navigation solutions able to provide the adequate performance required for such complex environments.

Within this framework, visual-based solutions, such as Visual Odometry (VO), Visual-Inertial Odometry (VIO) or Visual Simultaneous Localization and Mapping (VSLAM), are gaining ground because of their significant advantages: cameras are passive sensors immune to external interference, they are adaptable to all types of platforms and usually satisfy payload Size, Weight, and Power (SWaP) requirements. Visual-based techniques refer to the process of estimating the camera ego-motion based on the changes that motion induces on the images registered by a single or multiple cameras [5]. Recent surveys provides rigorous literature overviews of existing vision-based navigation solutions, which are generally classified into two categories: geometry-based and deep learning-based methods [2] [1] [3] [4]. However, although they have been an active research during several years and present promising results, real flight tests stand out the problems of using these techniques to estimate the relative movement: lack of adaptability to low-textured and/or light-challenging scenarios, loss of track during abrupt maneuvers due to small frame-to-frame overlaps or high sensitivity to weather effects such as fog, rain or night conditions [6] [5].

To overcome these limitations, two different novelties are being introduced into the problem by researchers: on the one hand, the application of Machine Learning (ML) techniques, which are becoming a major opportunity for the aviation industry due to their high adaptability to challenging scenarios. However, they conversely present a significant challenge due to the strict certification procedures in this industry [1] [4]. On the other hand, new hardware components are emerging and/or becoming more affordable for their general use. This is the case, for instance, of infrared cameras and all its variants (near-infrared, mid- and long-wave infrared) and the emerging event cameras. Thus, the combination of these cutting-edge technologies and the new (or even existing) techniques for data processing open up opportunities to solve the constraints related to existing visual-based navigation approaches.

Within this context, the present work focuses on bridging the gap related to the navigation in extremely low-illuminated or night-time scenarios using visual information, which to the

© 2023 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/ republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

extent of our knowledge, is still an open problem. To this end, the proposed solution builds upon one of the most powerful characteristics of event cameras: their wide dynamic range. Unlike traditional cameras, whose dynamic range usually is around 60 dB, the dynamic range of event cameras achieves more than 140 dB, allowing the users to obtain meaningful data even under extreme illuminations conditions [17]. The proposed algorithm is composed by two different threads running in parallel. One of them focused on generating a reference map which is only updated when new truthful information appears on the scene, while the other thread is in charge of estimating the relative displacement of the aircraft by matching real-time event-generated images against the reference map at a high frequency.

Although both threads share the same source of information, the events provided by the dynamic vision sensor, the type of images generated during both steps are essentially different, making traditional methods not feasible for their alignment. Then, another novelty introduced in this work corresponds to the application of Mutual Information (MI) theory to achieve a robust image alignment. MI has emerged in recent years as a powerful method for quantifying the statistical dependence between images [37]. It provides an effective criterion for measuring the similarity of images while not imposing any constraint on the modalities involved [32].

The rest of the paper is organized as follows. In Section II the State of the art related to the relevant topics is presented. The proposed algorithm is accurately described in Section III. The preliminary experiments and results are introduced in Section IV. Section V gathers the conclusions drawn along the algorithm development and test. Finally, the future work can be found in Section VI.

II. State of the art

This section presents a literature review related to the main topics introduced in this work: Visual navigation in low-illumination conditions, Event camera-based localization and Mutual Information applied to image alignment.

A. Visual navigation in low illumination conditions

Visual-based localization methods correspond to a widespread active research area since many years ago. Generally, the majority of approaches proposed so far tackle only favorable conditions: good illumination, static scene with enough texture in the images and sufficient scene overlap [5] [1] [3] [4]. However, recent studies have been conducted on analyzing different solutions for particularly challenging scenarios. Regarding low-illuminated or night-time environments, the proposed approaches can be categorized as software or hardware solutions, depending on how the problem is addressed.

Regarding software solutions two options are mainly conducted: image enhancement and camera control. With respect to the first one, some studies have focused on different methods for enhancing images quality under challenging night conditions. In [6], methods such as: adjusting the luminance of an image by applying Gamma Correction, removing the blue channel from the RGB image and convert it to grayscale by averaging the remaining red and green channels, converting the image from the traditional RGB color space into a different color space focused on representing lightness or intensity or building models to simulate parts of the human visual system, are presented. However, despite they are simple and effective, their application is limited in dark environments because the noise amplification they cause [6] [15].

The second software approach focuses on obtaining an optimal image through camera control. The idea is to capture the whole dynamic range of the scene by acquiring and combining several images taken at different exposure values. Then, the final image is obtained through multiple image alignment [7]. Unfortunately, the main drawback of these approaches corresponds to the difficulty of meeting real-time requirements. Nevertheless, recent advances in deep learning show an impressive performance in recovering the whole dynamic range from a single image via a deep neural network [16].

On the other hand, research has been also conducted to enable the navigation in low luminosity conditions by introducing different hardware components. In the case of unmanned ground vehicles, most of these approaches have either focused on the use of active sensors, such as Lidar, or introducing active illuminators. However, for the case of aerial vehicles this is only practical for very specific cases, such as the work presented in [14], which integrates onboard LEDs to provide illumination allowing to perform autonomous inspection of tunnels using aerial vehicles. In general, most studies involving flying platforms focus on using thermal-infrared cameras [12] [11] or combining thermal-infrared and visual cameras to achieve localization in different scenarios [13] [10] [8]. Finally, in [9] a new specific type of sensor is proposed, a new low-light sensor to provide both day and night imaging capability.

B. Event camera-based navigation

The emergence of event cameras shed light on overcoming some of the traditional limitations in existing visual-based navigation methods. The most complete reference for event-based vision corresponds to [17] which not only introduces the dynamic vision sensors from their working principle, but also presents a rigorous and comprehensive overview of researches according to the addressed task, such as feature detection and tracking, image segmentation or visual odometry.

For the specific case of navigation in low-illuminated conditions, some studies has been already conducted. For instance, [25] and [26] present a robust and accurate state estimator for High Dynamic Range (HDR) and high-speed scenarios by combining events, RGB images and inertial measurements. [18] also leverages outstanding properties of the event cameras proposing the exploration of lava tubes by a Mars helicopter by combining an event-based frontend with a filter-based backend for incorporating IMU measurements.

On the other hand, although not focused on challenging light conditions, some previous researches have worked not only on estimating the camera ego-motion, but also on reconstructing a map of the environment, similar to the method presented in this work. [21] proposes a method which can perform real-time 3D reconstruction based on three decoupled probabilistic filters: estimating the camera motion, scene logarithmic intensity gradient and scene inverse depth. [22] performs fast camera tracking while recovering a semidense 3D map of the environment based on image-to-model alignment. [24] introduces 3D contrast maximization instead of the commonly used Image of Warped Events (IWE) for motion tracking and mapping with an event camera, performing a joint optimization over motion and structure.

In recent years some researches have started to explore the application of Deep Learning (DL) techniques for event-based vision applications. In this line, [26] provides a comprehensive taxonomy for existing DL-based methods conducting benchmark experiments for these methods. The analyzed methods are grouped into two major categories: image reconstruction and restoration; and event-based scene understanding 3D vision.

Finally, an event-based stereo VIO system, namely ESVIO, has been presented which fuses events' depth, time-surface images, and pre-integrated inertial measurement to estimate the camera motion in a sliding-window non-linear optimization framework [20].

C. Mutual Information for image alignment

The robustness of Mutual Information (MI) against illumination changes and its independence on the modalities involved during the image processing, have positioned MI as a powerful tool to quantify the similarity of multi-modal imagery for their alignment. In this context, one of the most important and pioneer fields where MI is broadly used corresponds to the medical imaging domain. Numerous works have focused on registering Magnetic Resonance (MR) images with Computed Tomography (CT) images, but also on aligning complex 3D models to real scenes based on different MI formulations [29] [31] [36] [37]. In this field, some other studies can be found which specifically assess the accuracy of MI and compare its performance when applying different formulations as metrics, such as normalized mutual information and cross correlation [30] [35].

However, MI faces also some limitations, consequently, specific works can be found for dealing with these constrains. For instance, MI requires estimating the joint distribution of the two images, which not always converge fast and accurately. To overcome this inconvenience, [34] presents a non-parametric sampling method for obtaining the joint distribution of a pair of images resulting with lower bias and improved convergence rates. On the other hand, when applying MI, only the relationship between pixels' pairs is considered, losing much of the spatial information inherent in images. [32] presents an extension to MI called Regional Mutual Information (RMI) that also takes neighborhood regions of corresponding pixels into account.

Finally, in recent years, the growing availability of numerous imaging modalities in different sectors has made the application of MI suitable for many other different applications than the medical imagery. One of the most important one corresponds to image registration for Multi-Modal Remote Sensing (MMRS). In this topic, [28] presents a review of traditional and advanced methods proposed for MMRS image registration. Also, some works can be found on matching stereoscopic images applying MI by computing a global joint histogram from the stereo pair [33].

III. Proposed Solution

The present work pursues enabling the aerial navigation in low luminosity environments, including night-time, by exploiting one of the most powerful characteristics of dynamic vision sensors: their high dynamic range. The proposed solution entirely relies on the information provided by a unique event camera attached to a flying platform pointing downwards. These cameras, instead of capturing frames at a fixed rate, provide asynchronous information when brightness changes in any pixel of the camera field of view are detected. Then, every reported event contains: the specific pixel location where the illumination change was registered, its associated time and the sign of change, to indicate if the pixel becomes lighter or darker. To facilitate the processing of the captured events, the information contained in them is used to generate images, which are further processed.

The developed algorithm is composed by two threads running in parallel, as it is depicted in Fig. 1:

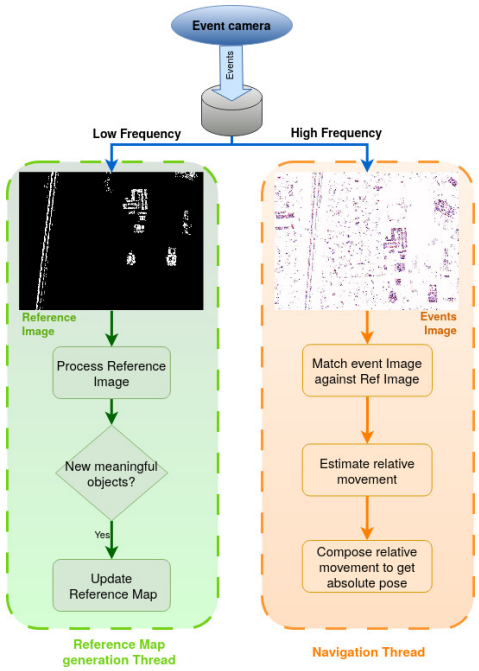

Fig. 1. Threads of the proposed solution.

- **Reference Map generation Thread**: It combines spatio-temporal information provided by the event camera to extract high-level features on the scene and reconstruct a 2D Reference Map based on them. This map is updated at a low frequency and only when new reliable shapes or objects appear on the scene.
- **Navigation Thread**: It matches real-time event-generated images against the Reference Map by applying Mutual Information. Once the alignment between the map and the image is completed, the relative displacement between them is computed. By concatenating all relative displacements, the whole trajectory relative to the initial point is achieved. This thread runs periodically at a high frequency.

The steps involved in both pipelines are further detailed in following subsections.

A. Reference Map generation Thread

Fig. 2 shows a day-time and a night-time image taken exactly at the same position by the same camera. Unlike the well-illuminated picture, the night-time image presents a poor texture and an absence of details, since most of the areas of the scene remain at the same level of darkness. Therefore, the first challenge of the algorithm corresponds to the identification of meaningful information in the night-time image.

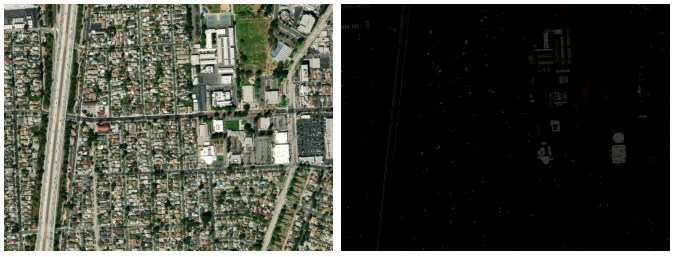

Fig. 2. Day (left) and night-time (right) aerial images taken at the same pose.

The Reference Map generation is performed in three steps: detection, extraction and tracking of a determined number of truthful objects in the scene considering their size and position. The main idea behind the map generation is to keep the biggest shapes while ensuring a proper distribution of them in the field of view of the camera. Other captured events which do not conform any of these shapes are removed from the scene. By removing small or noisy elements while keeping well-distributed large shapes, the trustworthiness of the subsequent matching process is increased. This process is repeated along the entire trajectory to ensure that a proper number of well-defined shapes are always identified and tracked. The diagram in Fig. 3 shows all steps involved in the Reference Map generation Thread.

0) Map Image available: At a rate of 1 HZ, the spatio-temporal data provided by the dynamic vision sensor is processed to generate a new Map Image.
1) Objects detection: The Map Image is processed to identify all objects on the scene as well as their size and position.

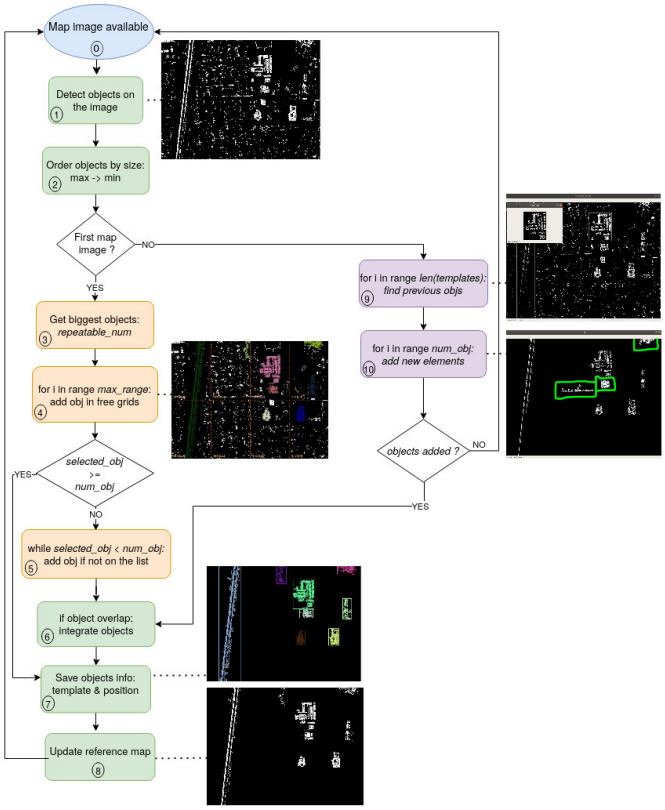

Fig. 3. Flow diagram of the Reference Map generation thread

2) Objects ordering: The detected objects are sorted out by their size, being the larger ones the firsts in the list.
3) Biggest objects selection: In order to ensure a well distribution of the selected shapes, the Map Image is divided into several grids of equal size. However, although the algorithm considers the object's position to ensure its homogeneous distribution, an introduced parameter allows to specify how many objects can repeat in the same grid. This is to avoid discarding large elements even if they are in the same region. Then, during this step, the biggest number of repeatable objects are selected and kept into the image, without considering their location.
4) Well-distributed objects selection: The first max_range objects stored in the sorted list, another configurable parameter, are analyzed to identify and keep those which are in currently empty grids.
5) Not Well-distributed objects selection: If not enough number of objects (determined by the parameter num_obj) have been gathered in the previous two steps, the next biggest objects that have not been analyzed yet are added to the Map, even if they are in already occupied grids. This ensures that a minimum number of objects are tracked.
6) Object overlap identification: As a simplification step, the contour of all selected objects is checked. If an overlap is found between two or more objects, they

are combined into a single object adapting the contour properly to contain all objects involved.

7) Save objects info: The information related to all selected objects is stored for the next iteration: template, position, and size.
8) Map Update: If it is the first iteration of the algorithm, the Reference Map is generated by keeping all objects selected during the previous steps. In the following iterations, the Reference Map is only updated if new relevant elements have been identified which are not present in the map.
9) Objects matching: After the first iteration, a map ob objects has been created. In the subsequent iterations, objects found in the new image are compared to those in the Map. The objective is to find and track each object in the Map with objects found in the new image.
10) Objects addition: If not all previous templates have been match because some of the objects are already out of the scene, the largest objects not previously matched are added to keep the same number of selected objects. Only if new objects have been added, the map is updated again.

B. Navigation Thread

While the Reference Map generation Thread is in charge of building the map of the environment along the trajectory, the Navigation Thread focuses on estimating the camera ego-motion using this map as reference. To this end, real-time event-generated images are composed at a high frequency for their subsequent alignment with the Reference Map. The alignment is done by applying Mutual Information, which provides a quantitative measure of similarity between images. Once the matching process is finished, the relative displacement between the Reference Map and the actual event-generated image is converted from pixels to meters by applying the Pinhole camera projection. Finally, by concatenating all computed relative displacements along the time, the whole trajectory referred to the initial point is easily extracted, similarly to VO/VIO solutions.

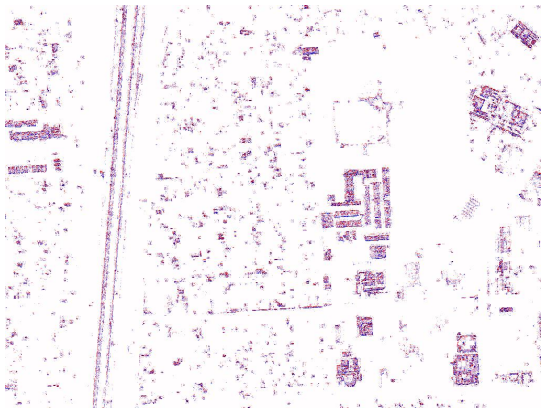

Fig. 4. Real-time event-generated image

1) **Image Alignment**: The image alignment or registration is achieved by computing the MI value between the Reference Map and the real-time event-generated images. The MI between two random variables quantifies their mutual statistical dependence, which is commonly used to guide multi-modal image alignment. MI criteria allows to quantify how much information one image contains about the other. Then, the images are geometrically aligned when the MI registers its maximum value, that is, when the amount of information they contain about each other is maximal.

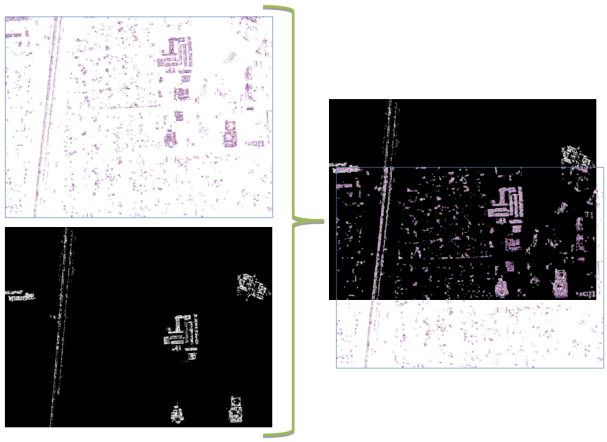

Fig. 5. MI-based match between the Ref Map and an event-generated image.

The MI of two images A and B can be expressed in terms of the marginal and joint entropies:

$$MI(A,B) = H(A) + H(B) - H(A,B) \tag{1}$$

The marginal entropy can be interpreted as the uncertainty related to a random variable, while the joint entropy as the uncertainty when combining two images. From Equation 1 it can be said that correct registration of the images is achieved by minimizing their joint entropy, in other words, the joint entropy is minimum when the images are totally aligned. The entropy of an image can be computed by estimating the probability distribution of its pixel intensities. The most common measure of entropy used in image processing corresponds to the Shannon-Wiener formulation:

$$H(A) = \sum_{a} -P_A(a) \log P_A(a) \tag{2}$$

$$H(B) = \sum_{b} -P_B(b) \log P_B(b) \tag{3}$$

$$H(A,B) = \sum_{a,b} -P_{A,B}(a,b) \log P_{A,B}(a,b) \tag{4}$$

The joint probability mass function $P_{A,B}(a,b)$ can be defined from joint histogram as:

$$P_{A,B}(a,b) = h(a,b) / \sum_{a,b} h(a,b) \tag{5}$$

where the joint histogram h is computed as:

$$h = \begin{bmatrix} h(0,0) & h(0,1) & ... & h(0,N-1) \\ h(1,0) & h(1,1) & ... & h(1,N-1) \\ : & : & ... & : \\ h(M-1,0) & h(M-1,1) & ... & h(M-1,N-1) \end{bmatrix} \tag{6}$$

In this equation, $h(a,b)$ corresponds to the number of point pairs for which the intensity in one image is a and for the other image is b, while M and N define the ranges of the intensity values of the pixels [28].

2) **Compute the relative displacement in real world**: Once the image alignment is completed, the next step corresponds to project the displacement in the image plane into displacements in the real world. This geometric transformation is achieved by inverting the Pinhole camera projection, illustrated in Figure 6.

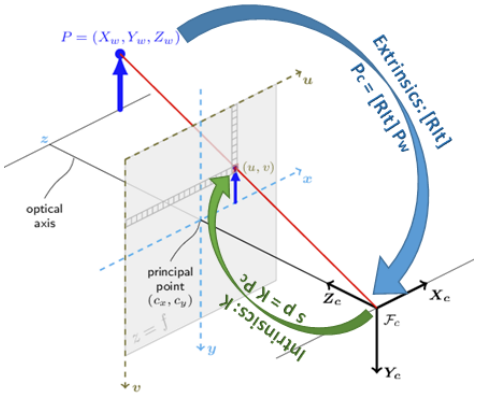

Fig. 6. Pinhole camera projection

The Pinhole camera projection allows to convert a 3D point expressed with respect to the world coordinate system, $P_W = (X_W, Y_W, Z_W)$, to a 2D pixel, $p = (u, v)$, in the image plane. As can be seen in Figure 6 this process is accomplished by two different steps. The first step is achieved by applying the homogeneous transformation, which is composed out of R, the 3-by-3 rotation matrix, and t, the 3-by-1 translation vector:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{7}$$

This homogeneous transformation allows to transform a 3D point in world coordinates, P_W, to a 3D point in camera coordinates, $P_C = (X_C, Y_C, Z_C)$, by applying the Extrinsics, i.e., the rotation (R) and translation (t) that describe the change of coordinates from world to camera coordinate systems. The contracted form of the Equation 7 is given by:

$$P_c = R\, P_w + t \tag{8}$$

Resolving for P_w, we have:

$$P_w = (P_c - t)R^{-1} \tag{9}$$

On the other hand, the projective transformation, enabled by the camera's intrinsic parameters, projects 3D points in the camera coordinate system, P_C, to 2D pixel coordinates, p. The camera intrinsic matrix, K, is composed of the focal lengths f_x and f_y, which are expressed in pixel units, and the principal point (c_x, c_y). The projective transformation is formulated then as:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \tag{10}$$

where s is the projective transformation's arbitrary scaling, not part of the camera model. The contracted form of the Equation 10 is given by:

$$s\, p = K\, P_c \tag{11}$$

Resolving for P_c, we have:

$$P_c = s\, p\, K^{-1} \tag{12}$$

Combining equations 9 and 12 we have the transformation that maps 2D points in the image plane, (u, v), to 3D points in world coordinates, P_W, which is the transformation required to estimate displacements in real world from the image alignment.

$$P_w = ((s\, p)K^{-1} - t)R^{-1} \tag{13}$$

3) **Compose whole trajectory** The initial pose of the camera establishes the origin of the World coordinate system, O_W, which is shared by both: Reference Map and localization updates. Since the relative displacements obtained by the two previous steps provide relative displacements in the current camera frame, P_{C_i}, they must be converted to the World reference system to obtain the whole trajectory relative to the initial point. The form of all rigid transforms is a rotation R followed by a translation t, which was previously introduced in Equation 8.

IV. Results

To validate the proposed algorithm, a simulator based on OSG-Earth is being used as the tool to obtain real satellite imagery for the designed trajectories. OSG-Earth corresponds to a 3D Earth surface rendering toolkit able to acquire satellite images based on specific camera poses. OSG-Earth-based simulator provides immense benefits: real images taken from real scenarios, configurable sensors, easily modifiable trajectories, and ground truth availability for analyzing the results.

Moreover, to compare the results, ORB-SLAM3 was used as the reference navigation algorithm. While the performance of ORB-SLAM3 showed exceptional results under day-time scenarios, it is only able to initialize and estimate the trajectory for 3 seconds (0.472% of the trajectory) before losing the track under night conditions. Figure 7 shows a screenshoot of the algorithm running in both situations. During the sunset and sunrise, i.e. low-illumination conditions, the position error

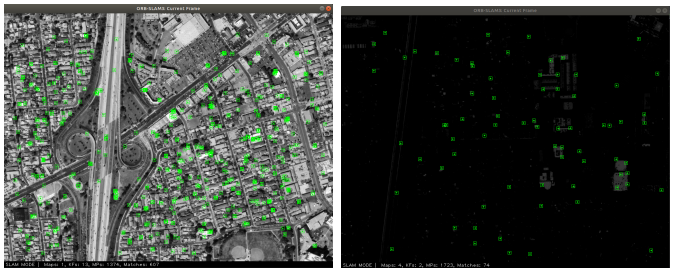

Fig. 7. ORB-SLAM solution applied to day (left) and night (right) trajectories.

increases between 4% and 26%, while the orientation error grows up to 52%.

The simulated dynamic vision sensor used for these experiments provides a resolution of 1024x768 pixels. All the tested trajectories were designed to last 250 seconds, which implies a mean of 4,672.83 meters of traveled distance. The flying altitude was established around 900 meters. The operation condition for all trials was totally night-time. An example of an image utilized during these experiments can be seen in Figure 2, on the right side. Finally, one assumption was made: the initial aircraft flying altitude above the ground was known.

The performance of the two implemented threads has been analyzed individually to fully understand the strengths and weaknesses of the proposed approach.

A. *Reference Map generation Thread Results*

As mentioned before, the Reference Map is not updated at a fixed rate, but only when new reliable shapes show up in the scene. Then, two factors have been analyzed: first, the accuracy of every individual update, and, then, the evolution of the generated map along time.

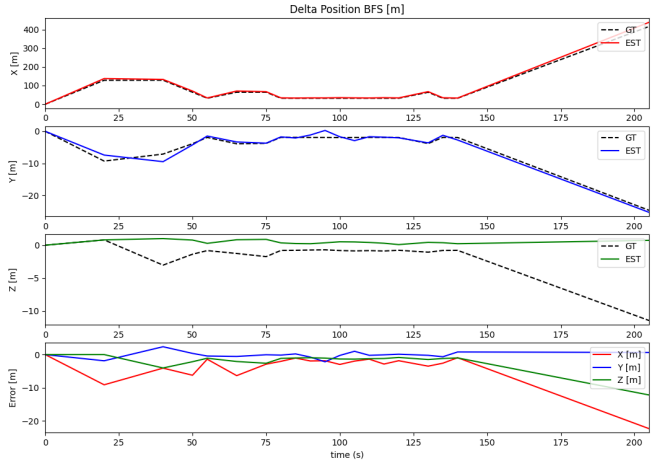

Fig. 8. Delta position between map updates along time.

Figure 8 depicts the delta position, i.e., changes in the camera position, between map updates. As we can see, the changes registered in the horizontal plane, axis X and Y, are accurately estimated, achieving a mean of 2.69 meters of error in this plane. However, movements in the vertical plane, Z, are not well estimated by the algorithm. Considering the distance traveled between map updates, displayed in Figure 9, the estimation error accumulated every traveled meter can be computed, which corresponds to 0.0508 meters, i.e., the 5.08% of the traveled distance.

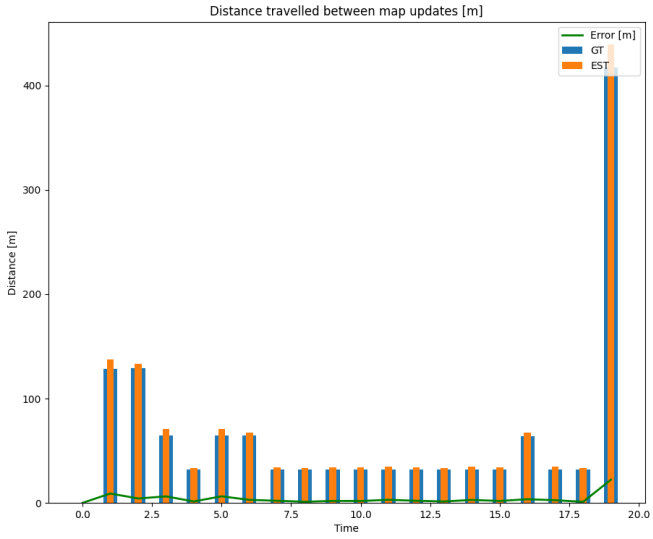

Fig. 9. Distance traveled between map updates along time.

On the other hand, Figure 10 shows the absolute position along time with respect to the initial pose. As expected, although the registered motion changes in the horizontal plane are estimated accurately, an accumulated drift over time is irremediable due to the relative nature of the proposed method. At the end of the tested trajectories, the X axis presents a mean of 77.4 meters of accumulated error, which divided by the traveled distance along this axis corresponds to 0.058 meters (5.8%), regarding the Y axis, the aggregated deviation stands around 1.987 meters, which corresponds to 0.025 meters by meter accomplished (2.5%), while the vertical axis accumulates almost a 70% of the error.

B. *Navigation Thread Results*

Real-time event-generated images are obtained and compared against the Reference Map at a fixed rate of $100Hz$, then, the relative position of the aircraft is periodically computed at this frequency as well. Although the Reference Map is used as a source of truth for estimating the camera ego-motion, we saw in the previous section that the generation of this map involves an inaccuracy which slowly grows over time. Thus, the localization step will target this drift as the best result, since no absolute references are available to remove it, i.e., the minimum possible error achieved by the Navigation Thread will correspond to the error accumulated during the map generation.

Figure 11 shows the localization error with respect to the three camera axis. This error corresponds to the specific mismatches when aligning real-time event-generated images and the Reference Map along time. Axis X presents an error

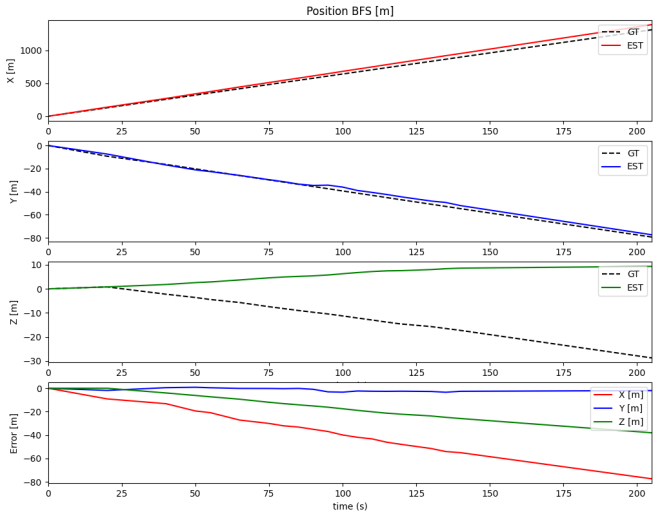

Fig. 10. Absolute position along time.

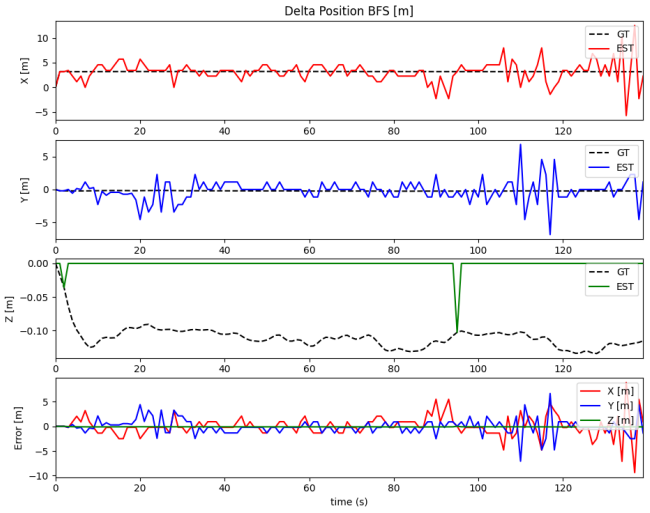

Fig. 11. Navigation error along the time.

mean of 1.34 meters, while its maximum value and variance corresponds to 9.38 and 2.56 meters, correspondingly. In the Y axis the mean error corresponds to 1.09, its maximum to 7.07 and its variance to 1.51. Regarding to the vertical axis, as mentioned before, changes in Z axis are not well estimated.

Finally, the whole trajectory in geodetic coordinates is displayed in the Figure 12. As expected, the estimated trajectory keeps very close to the ground truth trajectory during the initial stages of the flight, but diverges over time due to the accumulated error. Longer tests are required to be confirmed, but the performed experiments show that the accumulated drift of the whole algorithm almost coincides to the drift hauled by the Ref Map generation (5.08% of the distance traveled), i.e., not accumulated error is perceived in the Navigation Thread.

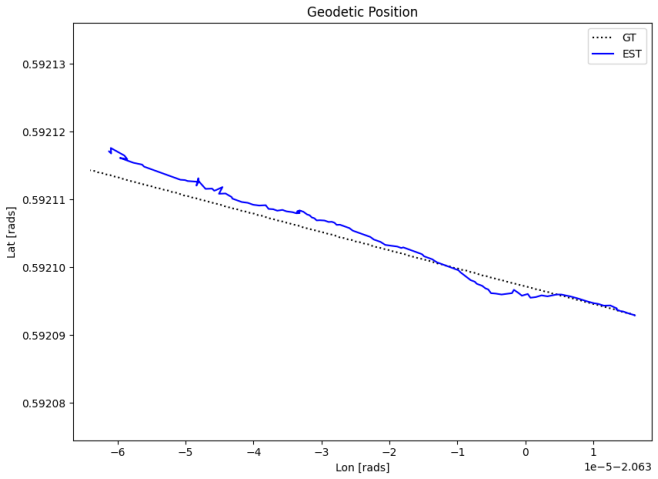

Fig. 12. Absolute position in geodetic coordinates

V. Conclusions

Existing visual-based localization solutions offer promising results under nominal scenarios. However, their overall performance highly degrades under challenging scenarios, which make them not reliable enough for real applications. In this work, a powerful property of dynamic vision sensors, their high dynamic range, is exploited to overcome the limitation faced in constrained illumination conditions, one of the biggest weaknesses of traditional methods.

The proposed solution not only pursues enabling the localization of an aerial vehicle based on an unique dynamic vision sensor attached to the platform pointing downwards, but also the reconstruction of 2-D map during the operation. Although all images involved in the algorithm are generated from events, different processing are applied to extract the required information. The introduction of Mutual Information as the criteria for image alignment is well suited for overcoming the difficulties of handling images of diverse nature, since this method makes no assumptions regarding the modalities involved.

Preliminary results show that, although improvements have been already identified (gathered in Section VI), the proposed solution can be the starting point for bridging the gap related to the aerial navigation in night-time scenarios using visual information, which to the extent of our knowledge is still an open problem.

Finally, as explained in Section IV, the performance of the proposed solution highly depends on the accuracy of the map generation, since it is during this step where most of the position drift is accumulated. Therefore, two reasons lead the map update to be done at low frequency: the scenario does not change in the short term, but more important, complex imaging process that cannot meet real-time requirements is performed to achieve the best results. Moreover, harboring the complexity in the Reference Map thread enables the localization step to be accomplished at high frequency, since simple processing is

required at this step.

VI. Next Steps

The presented method corresponds to the first version of the algorithm. Then, the main objective has been focused on demonstrating that the proposed approach is well-founded. Preliminary results show that the methodology is well-suited to enable the navigation in low-illuminated scenarios. Then, the following next steps will be implemented to improve and extend the performance of the solution:

- **Extend the solution to 6-DoF.** The current version of the algorithm is only able to estimate the displacement of the camera along the trajectory. The next version will incorporate not only the computation of changes in the position, but also in the orientation, achieving the whole camera pose estimation.
- **Addition of an IMU measurements.** In order to improve the relative position estimation and facilitate the attitude computation, the measurements of an Inertial Measurement Unit can be integrated for their fusion with the data provided by the event camera.
- **Variable time window for event processing** Real-time event-generated images are composed at a fixed rate. This is not optimal since in high dynamical situations enough events to achieve a reliable match are gathered very quickly, so there is no need to wait longer, but in some other circumstances, not sufficient events are available in the same time, implying a not accurate alignment. Since MI provides the uncertainty related to the matching process, a variable time window can be implemented, using the MI matching value to set the length of the time window at every individual step.
- **Remove accumulated drift.** The main drawback when developing relative localization methods corresponds to the drift accumulated over time, since it cannot be counteracted. A solution to be implemented corresponds to applying Image Registration for removing the drift periodically. To that end, pre-stored day-time georeferenced satellite images will be used to perform image registration between the reference map and the georeferenced images allowing to reset the accumulated error.

References

[1] Zhang, T., Hu, X., Xiao, J., Zhang, G. A survey of visual navigation: From geometry to embodied AI. Engineering Applications of Artificial Intelligence. Volume 114. 2022. 105036. ISSN 0952-1976.

[2] Qin, J., Li, M., Li, D., Zhong, J., Yang, K. 2022. A Survey on Visual Navigation and Positioning for Autonomous UUVs. 2022. Remote Sensing. 14. 3794. 10.3390/rs14153794.

[3] Delmerico, J., Scaramuzza, D. A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms for Flying Robots. IEEE International Conference on Robotics and Automation (ICRA). 2018, pp. 2502-2509, doi: 10.1109/ICRA.2018.8460664.

[4] Changhao, C.,Wang, B., Lu, C., Trigoni, N., Markham, A. A Survey on Deep Learning for Localization and Mapping: Towards the Age of Spatial Machine Intelligence. 2020.

[5] Scaramuzza, D., Fraundorfer, F. Visual Odometry: Tutorial. IEEE Robotics & Automation Magazine. Vol. 18. No. 4, p. 80-92. Dec. 2011, doi: 10.1109/MRA.2011.943233.

[6] Aladem, M., Baek, S. and Rawashdeh, S. Evaluation of Image Enhancement Techniques for Vision-Based Navigation under Low Illumination. Journal of Robotics. 2019. 1-15. 10.1155/2019/5015741.

[7] Kim, j., Jeon, h., Cho, Y., Kim, A. Dark Synthetic Vision: Lightweight Active Vision to Navigate in the Dark. IEEE Robotics and Automation Letters, vol. 6, no. 1, pp. 143-150, Jan. 2021, doi: 10.1109/LRA.2020.3035137.

[8] Fan, Y., Ding, M., Cao, Y. Vision algorithms for fixed-wing unmanned aerial vehicle landing system. Science China Technological Sciences. 2017. 60. 10.1007/s11431-016-0618-3.

[9] Chiu, H., Samarasekera, S., Mithun, N., Rajvanshi, A., Kaighn, K., Murray, G., Oskiper, T., Sizintsev, M., Kumar, R. Night-Time GPS-Denied Navigation and Situational Understanding Using Vision-Enhanced Low-Light Imager. Joint Navigation Conference. June 12-15, 2023.

[10] Bhatnagar, G., Wu, Q. M. J., Raman, B. Navigation and surveillance using night vision and image fusion. 2011 IEEE Symposium on Industrial Electronics and Applications, Langkawi, Malaysia, 2011, pp. 342-347, doi: 10.1109/ISIEA.2011.6108728.

[11] Daftry, S., Das, M., Delaune, J., Sorice, C., Hewitt, R., Reddy, S., Lytle, D., Gu, E., Matthies, L. Robust Vision-based Autonomous Navigation, Mapping and Landing for MAVs at Night. 2018

[12] Borges, P., Vidas, S. Practical Infrared Visual Odometry. IEEE Transactions on Intelligent Transportation Systems. 17. 1-9. 2016. 10.1109/TITS.2016.2515625.

[13] Brunner, C., Peynot, T., Vidal-Calleja, T., Underwood, J. Selective Combination of Visual and Thermal Imaging for Resilient Localization in Adverse Conditions: Day and Night, Smoke and Fire. Journal of Field Robotics. 30. 2013. 10.1002/rob.21464.

[14] Özaslan, T. et al. Autonomous Navigation and Mapping for Inspection of Penstocks and Tunnels With MAVs. IEEE Robotics and Automation Letters. Vol. 2. No. 3, pp. 1740-1747, July 2017, doi: 10.1109/LRA.2017.2699790.

[15] Shim, I., Lee, J.-Y., Kweon, I. S. Auto-adjusting camera exposure for outdoor robotics using gradient information. Proc. IEEE/RSJ Intl. Conf. Intell. Robots Syst., pp. 1011-1017, 2014.

[16] Liu, Y.-L. et al. Single-image HDR reconstruction by learning to reverse the camera pipeline. Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1651-1660, 2020.

[17] Gallego, G. et al. Event-Based Vision: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 44, pp. 154-180. 1 Jan. 2022, doi: 10.1109/TPAMI.2020.3008413.

[18] Mahlknecht, F., Gehrig, D., Nash, J., Rockenbauer, F., Morrell, B., Delaune, J., Scaramuzza, D. Exploring Event Camera-based Odometry for Planetary Robots. 2022.

[19] Milford, M., Kim, H., Leutenegger, S., Davison, A.J. Towards Visual SLAM with Event-based Cameras. 2015

[20] Liu, Z., Shi, D., Li, R., Yang, S. ESVIO: Event-Based Stereo Visual-Inertial Odometry. Sensors 2023, https://doi.org/10.3390/s23041998.

[21] Kim, H., Leutenegger, S., Davison, A. Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera. 2016. 349-364. 10.1007/978-3-319-46466-4-21.

[22] Rebecq, H., Horstschaefer, T., Gallego G., Scaramuzza, D. EVO: A Geometric Approach to Event-Based 6-DOF Parallel Tracking and Mapping in Real Time. IEEE Robotics and Automation Letters. vol. 2. no. 2. pp. 593-600. April 2017. doi: 10.1109/LRA.2016.2645143.

[23] Rebecq, H., Horstschaefer, T., Scaramuzza, D. Real-time Visual-Inertial Odometry for Event Cameras using Keyframe-based Nonlinear Optimization. 2017. 10.5244/C.31.16.

[24] Wang, Y., Yang, J., Peng, X., Wu, P., Gao, L., Huang, K., Chen, J., Kneip, L. Visual Odometry with an Event Camera Using Continuous Ray Warping and Volumetric Contrast Maximization. 2022. Sensors. 22(15), 5687. https://doi.org/10.3390/s22155687.

[25] Vidal, A.R., Rebecq, H., Horstschaefer, T., Scaramuzza, D. Hybrid, Frame and Event based Visual Inertial Odometry for Robust, Autonomous Navigation of Quadrotors. 201). abs/1709.06310.

[26] Vidal, A.R., Rebecq, H., Horstschaefer, T., Scaramuzza, D. Ultimate SLAM: Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High-Speed Scenarios. IEEE Robotics and Automation Letters. 2017. 994-1001.

[27] Zheng, X., Ye-Peng Liu, Y.L., Tongyan Hua, T.P., Weiming Zhang, D.T., Wang, L. Deep Learning for Event-based Vision: A Comprehensive Survey and Benchmarks. 2023. ArXiv abs/2302.08890 2023

[28] Zhang, X., Leng, C., Hong, Y., Pei, Z., Cheng, I., Basu, A. Multimodal Remote Sensing Image Registration Methods and Advancements: A Survey. Remote Sens. 2021, 13, 5128. https://doi.org/10.3390/rs13245128

[29] Viola, P., Wells III, W.M. Alignment by Maximization of Mutual Information. International Journal of Computer Vision 24, 137–154 (1997). https://doi.org/10.1023/A:1007958904918

[30] Roshini, V.S., Revathy, K. Using mutual information and cross correlation as metrics for registration of images. 2008. Journal of Theoretical and Applied Information. 4.

[31] Sruthi, M. V., Shree, U., Soundararajan, K. Enhanced Multimodality Image Registration Based On Mutual Information. IOSR Journal of VLSI and Signal Processing (IOSR-JVSP). Volume 6, Issue 1, Ver. I (Jan. -Feb. 2016), PP 09-12. DOI: 10.9790/4200-06110912

[32] Russakoff, D.B., Tomasi, C., Rohlfing, T., Maurer, C.R.. Image Similarity Using Mutual Information of Regions. Computer Vision - ECCV 2004. Springer. https://doi.org/10.1007/978-3-540-24672-5.

[33] Fookes, C., Bennamoun, M., Lamanna, A. Improved stereo image matching using mutual information and hierarchical prior probabilities. 2002. International Conference on Pattern Recognition. pp. 937-940 vol.2, doi: 10.1109/ICPR.2002.1048459.

[34] Dowson, N., Bowden, R., Kadir, T. Image template matching using Mutual Information and NP-Windows. International Conference on Pattern Recognition (ICPR). 2006. DOI:10.1109/ICPR.2006.691.

[35] Pluim, J.P.W, Maintz, J.B.A, Viergever, M.A. Mutual information matching in multiresolution contexts. Image and Vision Computing. Volume 19, Issues 1–2. 2001. Pages 45-52. ISSN 0262-8856. https://doi.org/10.1016/S0262-8856(00)00054-8.

[36] Kosiński, W., Michalak, P., Gut, P. Robust Image Registration Based on Mutual Information Measure. Journal of Signal and Information Processing. Vol. 3 No. 2. 2012. pp. 175-178. doi: 10.4236/jsip.2012.32023.

[37] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P. Multi-modality image registration by maximization of mutualinformation. IEEE Trans. Med. Imaging. 16. 1996.14-22. 10.1109/MMBIA.1996.534053.

https://dspace.org/

Enabling UAVs night-time navigation through mutual information-based matching of event-generated images

Escudero, Naiara; Hardt, Michael W.; Inalhan, Gokhan

2023-11-10
Institute of Electrical and Electronics Engineers (IEEE)
https://doi.org/10.1109/DASC58513.2023.10311251
Escudero N, Hardt MW, Inalhan G. (2023) Enabling UAVs night-time navigation through mutual information-based matching of event-generated images. In: IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC) 2023, 1-5 October 2023, Barcelona, Spain

Downloaded from a DSpace repository. https://dspace.org/