A Neuromorphic Vision-Based Measurement for Robust Relative Localization in Future Space Exploration Missions

Mohammed Salah, Mohammed Chehadah, *Member, IEEE,* Muhammad Humais, Mohammed Wahbah, Abdulla Ayyad, *Member, IEEE,* Rana Azzam, Lakmal Seneviratne, and Yahya Zweiri, *Member, IEEE*

Abstract—Space exploration has witnessed revolutionary changes upon landing of the Perseverance Rover on the Martian surface and demonstrating the first flight beyond Earth by the Mars helicopter, Ingenuity. During their mission on Mars, Perseverance Rover and Ingenuity collaboratively explore the Martian surface, where Ingenuity scouts terrain information for rover's safe traversability. Hence, determining the relative poses between both the platforms is of paramount importance for the success of this mission. Driven by this necessity, this work proposes a robust relative localization system based on a fusion of neuromorphic vision-based measurements (NVBMs) and inertial measurements. The emergence of neuromorphic vision triggered a paradigm shift in the computer vision community, due to its unique working principle delineated with asynchronous events triggered by variations of light intensities occurring in the scene. This implies that observations cannot be acquired in static scenes due to illumination invariance. To circumvent this limitation, high frequency active landmarks are inserted in the scene to guarantee consistent event firing. These landmarks are adopted as salient features to facilitate relative localization. A novel event-based landmark identification algorithm using Gaussian Mixture Models (GMM) is developed for matching the landmarks correspondences formulating our NVBMs. The NVBMs are fused with inertial measurements in proposed state estimators, landmark tracking Kalman filter (LTKF) and translation decoupled Kalman filter (TDKF) for landmark tracking and relative localization, respectively. The proposed system was tested in a variety of experiments and has outperformed state-of-the-art approaches in accuracy and range.

Index Terms—Gaussian mixture models (GMM), flickering landmarks, Neuromorphic vision-based measurements (NVBM), landmark tracking Kalman filter (LTKF), translation decoupled Kalman filter (TDKF), space robotics

I. INTRODUCTION

S PACE exploration has delivered a wealth of scientific discoveries on the history and evolution of the solar system. Space exploration recently witnessed a paradigm shift as space agencies started to deploy unmanned platforms for various space missions allowing exploration of areas that were once inaccessible to mankind. NASA marked a major milestone deploying the first rover-copter team comprised of the Mars helicopter, Ingenuity, and the Mars rover, Perseverance, on Mars [1]. The Ingenuity scouts the terrain to facilitate terrain mapping and boost the rover's navigation performance. However, relative localization between both platforms is of fundamental importance for their safe navigation.

Traditional relative localization systems relying on radio-frequency identification (RFID) and ultra-wideband (UWB) do not provide sufficient relative localization accuracy [2], [3]. On the other hand, vision-based measurements (VBMs) using conventional imaging sensors tend to deteriorate in harsh space environments [4]. To alleviate the shortcomings of such systems such as low dynamic range and high latency, there is a surging demand for new relative localization paradigms based on robust sensing modalities that could endure harsh environmental conditions and hence enable persistent relative localization.

Several approaches have been proposed for relative localization systems, mainly relying on radio frequency identification (RFID), ultrawideband (UWB) measurements, VBM, and neuromorphic vision-based measurements (NVBM). RFID-based measurements have been generally utilized for indoor localization systems due to their passive nature [5]–[8]. It was not until recently, though, that RFID-based relative localization systems were developed due to the large tracking errors of onboard antennae. Li et al. [7] proposed a relative localization system that relies on phase, received signal strength indicator (RSSI), and readability that is tolerant to the antenna tracking errors. However, such relative localization systems are impractical for space applications since RFID measurements require a fixed infrastructure for the tags.

UWB-based approaches are considered as one of the earliest approaches to relative localization, due to the low cost and ease of deployment of UWB. Previously, UWB measurements were utilized for localization in GPS-denied environments where fixed nodes were installed as anchors for positioning the target platform [9], [10]. Similar to RFID, such localization systems are difficult to utilize for space applications for the fixed infrastructure requirement. To overcome this challenge, Strader et al. [11] proposed a relative localization system that estimates the relative pose between two UAVs in-motion,

M. Salah, M. Chehadah, M. Humais, M. Wahbah, R. Azzam, L. Senevirante, and Y. Zweiri are with the Center for Autonomous Robotic Systems, Khalifa University, Abu Dhabi, United Arab Emirates. A. Ayyad is with Advanced Research and Innovation Center (ARIC), Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates. Also, L. Seneviratne is the director of Center for Autonomous Robotic Systems, and Y. Zweiri is with the Department of Aerospace Engineering, both at Khalifa University, Abu Dhabi, United Arab Emirates.

augmenting range and displacement measurements from UWB sensor and a generic navigation system using nonlinear least squares. However, the approach was only tested in simulation and needs experimentation to validate robustness. Guler et al. [12] proposed a different approach based on Monte Carlo localization (MCL) sequential methods for modelling the motion of the tag quadrotor. The quadrotor's motion was restricted to the horizantal plane where it was flying at constant altitude.

Landmark-based positioning systems proved to be superior to UWB and RFID in terms of localization accuracy. In such systems, conventional imaging sensors extract landmark positions as VBM to reconstruct the agent pose [13]–[16]. Walter et al. [17] proposed a relative localization system for quadrotors relying on a conventional camera with ultraviolet pass filter and ultraviolet LEDs emitting light at frequencies much lower than the visible light spectrum. Similarly, Cutler et al. [18] utilized an infrared camera with infrared emitting landmarks. However, the major drawback of these systems is the degraded performance of the conventional imaging sensors under extreme space conditions due to their high latency and low dynamic range. Hence, utilizing a different imaging pipeline is substantial for the robustness of such systems in uncertain space environments.

Neuromorphic vision sensors (NVS) provide improved performance in extreme environments given their high dynamic range and low latency achieved by virtue of their asynchronous nature [19]–[21], which motivate the move towards neuromorphic vision technologies. These unprecedented capabilities of the NVS have been utilized in space [4], [22], tactile sensing [23]–[25], and autonomous navigation applications [26], [27]. More importantly, the NVS is currently utilized as an imaging sensor for various vision-based localization systems. Mueggler et al. [28] and Valeiras et al. [29] show the superiority of the NVS over conventional imaging sensors for UAV navigation as traditional cameras suffer from motion blur and degraded feature extraction from the observed scenes. Nguyen et al. [30] proposed a stacked spatial LSTM network to robustly extract the features observed from accumulated event frames and learn their spatial dependencies in the image feature space. Yuan and Ramalingam [31] developed a fast spatio-temporal binning technique to detect lines that correspond to edges in the observed scene as features for localization. In all of these approaches, determining the relative motion between the NVS and the observed scene is a must for events to be generated, posing a major limitation in absence of activity in the environment. To alleviate this restriction, flickering LED luminaries are used as a constant source of intensity variations to ensure consistent event generation and as features for localization [32]–[35], irrespective of the scene dynamics. Chen et al. [32] proposed an indoor localization system using an NVS and flickering landmarks where the landmarks are characterized by unique frequency IDs without introducing complex data association algorithms. The drawback of these systems is the high levels of noise present in frequency detections leading to false landmarks identifications and large positioning errors. In addition, due to the sensor's low resolution, the positioning accuracy of these systems, 0.03 m within 1 meter range, which is insufficient for the targeted space scenario.

Previously developed relative localization systems suffer in terms of installation feasibility in space and their robustness in extreme environments. On the other hand, neuromorphic vision along with flickering landmarks demonstrates an improved system architecture for landmark-based positioning systems. However, robust landmark identification and localization accuracy remain a gap that needs to be addressed accordingly.

NVSs have a huge potential for future space missions given their unprecedented capabilities in harsh environmental conditions. Current space applications of the NVS are mainly limited to astronomical object tracking, such as stars and space debris [22], [36] and visual odometry for planetary robots [4]. In this work, we use NVS to enable relative localization under harsh space conditions. In our approach, neurmorphic vision-based measurements (NVBM) are fused with inertial measurements in robust state estimators providing state-of-the-art (SOTA) relative localization performance.

A. Contributions

The motivation of this article is to develop a reliable relative localization system for the Ingenuity and Perseverance Rover providing robust landmark identification and relative localization performance. Accordingly, this article comprises the following contributions:

1) A novel event-based landmark identification system is proposed to detect and identify salient flickering features by means of a frequency-based unsupervised clustering using Gaussian Mixture Models (GMMs) with a maximum frequency detection error of 3.21 Hz.
2) A measurement fusion of NVBMs and inertial measurements is developed using a landmark tracking Kalman filter (LTKF) and a translation decoupled Kalman filter (TDKF) for landmark tracking and relative localization.
3) The developed relative localization system was demonstrated through a space exploration scenario outperforming state-of-the-art approaches in terms of accuracy with a mean positioning error of 0.0052 m.

B. Structure of The Article

The rest of the article is organized as follows. Section II provides the notation followed in the paper and gives an overview of the proposed relative localization system. Section III discusses the theory behind the NVBMs formulation. Section IV illustrates the proposed LTKF and TDKF state estimators for landmark tracking and relative localization, respectively. Experimental validation of the developed system is provided in section V and section VI presents conclusions and future aspects of the developed work.

II. PRELIMINARIES

A. Notations and Reference Frames

The reference frames that are considered for developing our approach are shown in Fig. 2. An inertial frame $\mathcal{F}_I$ is defined of basis $[i_x, i_y, i_z]$ with i_z antiparallel to the gravity vector, and the IMU body fixed reference frame $\mathcal{F}_B$.
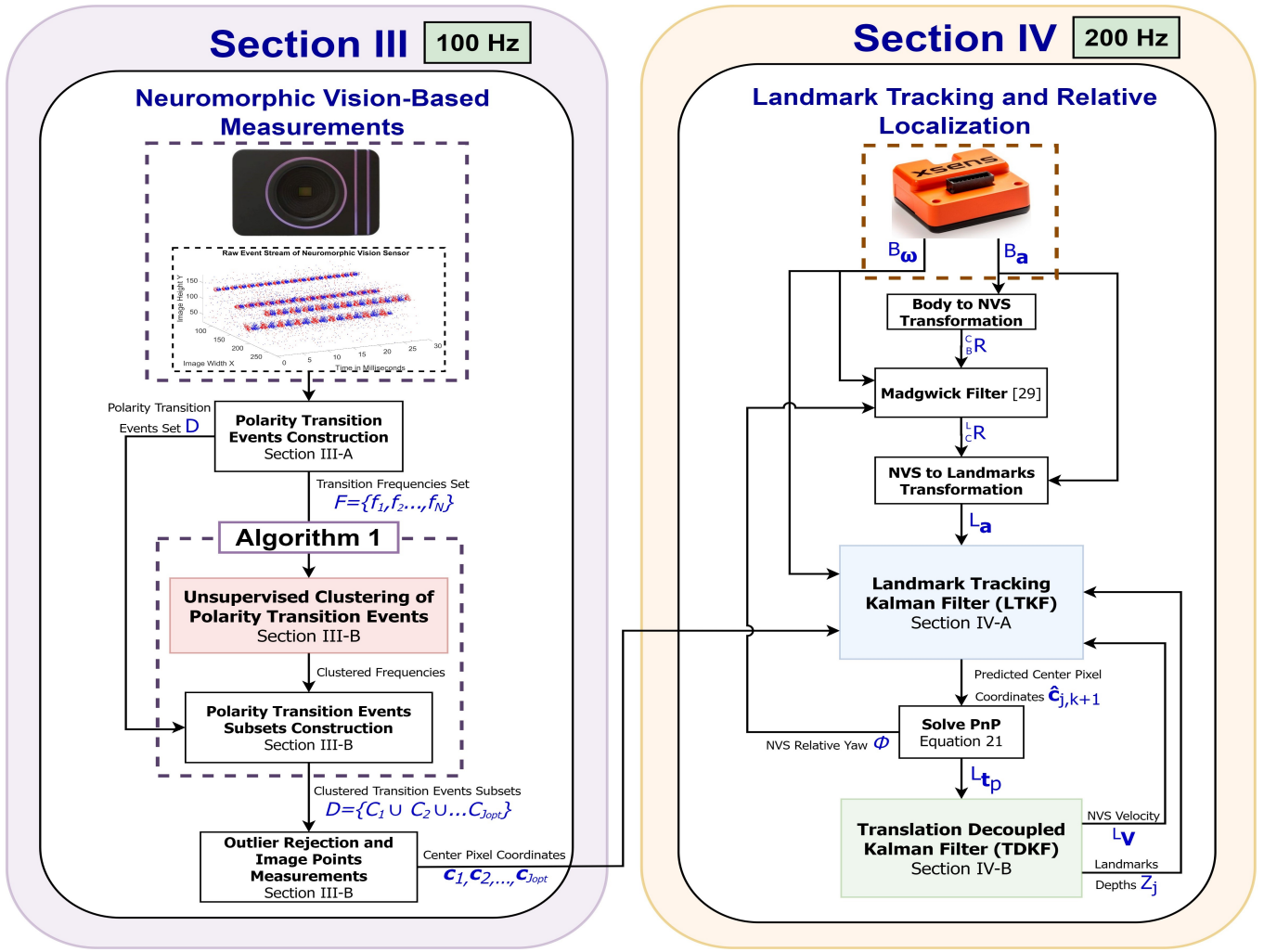
Fig. 1: Block diagram of the proposed relative localization system. The set of transition frequencies are clustered using GMMs and the landmarks pixel centers are extracted formulating the NVBMs at 100 Hz. The NVBMs are fed to the LTKF for tracking and relative localization is attained by the TDKF at 200 Hz.

Also, the landmarks reference frame $\mathcal{F}_L$ defining their relative positions with basis $[l_x, l_y, l_z]$, where l_z is parallel to i_z. We also define the NVS frame of reference $\mathcal{F}_C$, with rotation matrix $_B^C R = [b_x, b_y, b_z] \in \mathrm{SO}(3)$, giving the transformation from $\mathcal{F}_B$ to $\mathcal{F}_C$. A vector in a particular reference frame is denoted with a pre-superscript, e.g. $^I t$ is the translation vector expressed in the inertial frame. In addition, components of vectors are denoted with subscripts (i.e. $^I t = [^I t_x\, ^I t_y\, ^I t_z]^T$).

B. System Overview

A block diagram of the proposed system is shown in Fig. 1 relying on measurement streams from the NVS and IMU. The system comprises of two major subsystems for landmark identification and tracking, and relative localization, respectively. Transition frequencies from the raw event stream are constructed since they provide a better representation of the landmarks carrier waves. The set of transition frequencies are clustered using GMMs to extract the true frequencies of the flickering landmarks as identifiers due to the high levels of Gaussian noise accompanied with the transition frequency
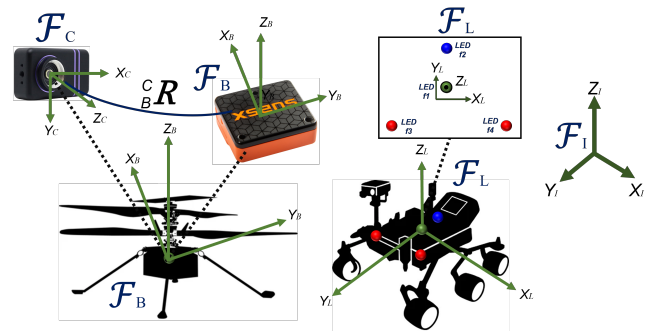

Fig. 2: The major frames of reference for the relative localization system.

measurements. Then, polarity transition events subsets are formulated given the frequency-based cluster assignments where blob detection is utilized for outlier rejection and extracting the landmarks 2D center pixel coordinates on each all events of each cluster formulating our NVBMs. One key aspect of the proposed system is that the landmarks flickering frequencies

were selected to be greater than 200 Hz to avoid correlation with low frequency background clutter. Accordingly, 100 Hz was selected as the frequency detection resolution to reliably detect all the landmarks frequencies as well as rejecting any frequency below 100 Hz. Nevertheless, we utilize the LTKF for tracking the landmarks based on their estimated pixel velocities running at the IMU higher update rate (200 Hz) providing robust tracking even when harsh maneuvers are performed.

It is important to note that to estimate pixel velocities, the Ingenuity linear and angular velocities along with the landmarks depths in the $\mathcal{F}_C$ need to be given as a priori. Therefore, we propose the TDKF running at also 200 Hz fusing relative pose estimates from a solution to the perspective-n-point (PnP) problem $^L\mathbf{t}_p$ and acceleration measurements $^B\mathbf{a}$ received from the IMU to estimate the Ingenuity linear velocity and its relative pose. We assume that the gyroscope bias is negligible and the angular velocity measurements are utilized directly in the LTKF. Surely, the IMU acceleration measurements need to be transformed to $\mathcal{F}_L$, where the transformation is obtained by the Madgwick filter and yaw angle estimates are utilized from the solve PnP block to account for yaw angle drift. We apply the proposed system on relative localization between the Ingenuity and Perseverance Rover, where $\mathcal{F}_B$ also resembles the Ingenuity body-fixed reference frame and $\mathcal{F}_L$ is aligned with rover body frame of reference as shown in Fig. 2.

III. NEURMORPHIC VISION-BASED MEASUREMENTS

A. Flickering Landmarks Frequency Detections

The NVS consists of an array of pixels each consisting of a photodiode, generating photocurrents stimulated by arriving photons, which get processed asyncronously and independently for every pixel. Every pixel circuit produces a binary output (i.e. polarity) corresponding to the log intensity changes. These binary outputs get spatial and temporal stamp producing an event

$$e_k \doteq (\boldsymbol{x}_k, t_k, p_k) \tag{1}$$

with $\boldsymbol{x} \in W \times H$ being pixel location where $W \in \{1, ..., w\}$ and $H \in \{1, ..., h\}$ representing an image of h pixels in height and w pixels in width, t is the timestamp, and $p \in \{-1, +1\}$ is the polarity of the event. Let a *polarity transition event*, Δ_k, be defined as:

$$\Delta_k \doteq e_k - e_l : \boldsymbol{x}_k \equiv \boldsymbol{x}_l \wedge t_k - t_l < \tau/2 \wedge p_k - p_l = 2 \tag{2}$$

where the difference operation $e_k - e_l$ is defined as a substraction of the timestamp and polarity elements (i.e. $\boldsymbol{x}_k$ is the same for Δ_k, e_k and e_l), and τ represents the maximum period that is considered for conecutive events. Note that it is assumed that $t_k > t_l$. The *polarity transition frequency* is defined as $f_k \doteq (2\Delta_k)_2^{-1}$, where the notation $(.)_i$ is used to reference the ith element in the tuple. The polarity transition events Δ and frequencies f are consecutively added to the sets D, $\{\Delta_k\} \cup D$, and F, $\{f_k\} \cup F$, respectively. The sets D and F store all Δ and f happened in the previous τ seconds. Hence $f_d = \tau^{-1}$ is called the *frequency detection resolution* which we select, in this work, to be 100 Hz (i.e. $\tau = 10$ ms) as
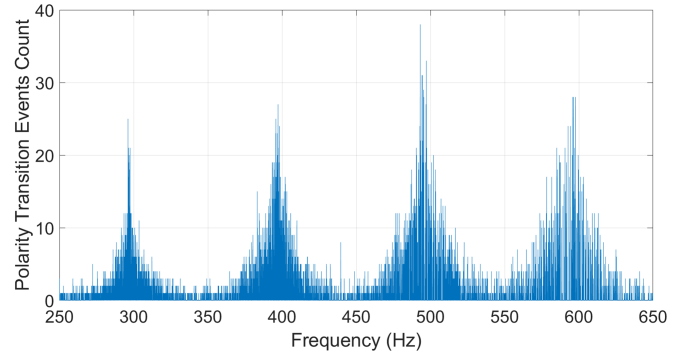

Fig. 3: A plot of the transition frequencies histogram following a Gaussian mixture model.

described in section II-B. As a consequence, polarity transition events with frequencies less than f_d cannot be detected.

The set F is the input to Algorithm 1, and the output is $C_1 \cup C_2 \cup ... \cup C_{J_{opt}} = D$ clustered sets with $j \in \{1, 2, ..., J_{opt}\}$. The optimal number of GMM clusters J_{opt} corresponding to the number of observed flickering landmarks is selected by minimizing the Bayesian information criterion (BIC) using the method detailed in the following Section and summarized in Algorithm 1.

B. Unsupervised clustering of polarity transition events

The clustering of polarity transition events represented in the set D is performed using the set of transition frequencies in F, those we ignore any spatial information about the polarity transition events during clustering.

Algorithm 1 specifies the procedure of clustering the transition frequencies in a set F for landmark identification. If we plot a histogram of F comprising of 172 transition frequency elements for four observed flickering landmarks with frequencies 300, 400, 500, and 600 Hz, shown in Fig. 3, it is obvious that the probability distribution of the transition frequencies $p(f_k)$ is dominated by a mixture of Gaussians, see Fig. 3. Hence, $p(f_k)$ is parameterized by the mixture components' mean μ_j, covariance Σ_j, mixture proportion π_j, and is modelled by

$$p(f_k) = \sum_{j=1}^{J} \mathcal{N}(f_k | \mu_j, \Sigma_j) \pi_j \tag{3}$$

where J represents the current number of clusters corresponding to the number of observed landmarks by the NVS.

The optimal values of μ_j, π_j, and Σ_j are the ones minimizing the $p(f_k)$ expected likelihood L [37] defined as

$$L = \sum_{n}^{N} \sum_{j} z_n^j \log \pi_j - \sum_{n}^{N} \sum_{j}^{J} z_n^j \frac{1}{2\Sigma}(f_n - \mu_j)^2 \tag{4}$$

using the expectation maximization (EM) algorithm, where N represents the number of elements in F.

The EM algorithm comprises of two iterative computations that are repeated until convergence: (1) the expectation (E)

step and (2) maximization (M) step. The E step evaluates the responsibilities $\beta(z_{nj})$ defined as

$$\beta(z_{nj}) = \frac{\pi_j \mathcal{N}(f_k|\mu_j, \Sigma_j)}{\sum_j \pi_j \mathcal{N}(f_k|\mu_j, \Sigma_j)}. \tag{5}$$

On the other hand, the M step revaluates μ_j, π_j, and Σ_j by

$$\mu_j = \frac{\sum_n \beta(z_{nj}) f_k}{\sum_n \beta(z_{nj})}, \tag{6}$$

$$\Sigma_j = \frac{1}{\sum_n \beta(z_{nj})} \sum_n \beta(z_{nj})(f_n - \mu_j)(f_n - \mu_j)^T, \tag{7}$$

$$\pi_j = \frac{\sum_n \beta(z_{nj})}{N}. \tag{8}$$

It is important to note that the EM algorithm requires J to be known as a priori which is not practical in the targeted application given the limited FOV of the NVS. Therefore, J_{opt} is selected based on minimizing the BIC γ expressed as [37]

$$\gamma = N \cdot \ln(\hat{\Sigma}_e^j) + \alpha \cdot \ln(N), \tag{9}$$

where α is 4 corresponding to the number of parameters to be estimated, μ_j, Σ_j, π_j, and J, and $\hat{\Sigma}_e^j$ is the error covariance defined as

$$\hat{\Sigma}_e^j = \frac{1}{N} \sum_j \sum_n (f_n - \mu_j). \tag{10}$$

The γ is evaluated on cluster hypotheses i ranging from 1 to J_{max}, which we select as 10 to maintain the algorithm's computational efficiency.

Once the polarity transition events have been assigned to clusters, every element f_n is associated with the corresponding element Δ_k. With such association, it is possible to filter out events that are not in the spatial neighborhood of the majority of the events in the cluster, and we achieve that by performing blob detection followed by masking the largest blob resulting in a binary image M_j [38], where $j \in \{1, ..., J_{opt}\}$. The moment of each image M_j is used to find the center pixel coordinates, $c_j = [u \ v]^T$, corresponding to each flickering landmark.

IV. LANDMARK TRACKING AND RELATIVE LOCALIZATION

A. Landmark Tracking Kalman Filter

The center pixel coordinates c_j corresponding to flickering landmarks in the image plane got estimated at the rate f_d, but it is possible to increase the rate of estimation to f_i corresponding to the NVS inertial states estimation rate. The center pixel velocities $\dot{c}_j$ can be estimated through the knowledge of the NVS inertial states and the relative location of the flickering landmarks. This requires the assumption that the landmarks are quasi static. The number of LTKFs used is J_{opt}. When a flickering landmark goes out of the FOV, its corresponding LTKF gets removed, and it get initialized again only once the corresponding flickering landmarks gets in the FOV again.

Algorithm 1 Unsupervised Clustering of Polarity Transition Events

1: **Inputs:**
    Set $F = \{f_1, f_2, ..., f_N\}$
2: **Outputs**: Clustered sets: $D = C_1 \cup C_2 \cup ... \cup C_{J_{opt}}$
3: **Initialize:**
    $p(f_k) = \sum_{j=1}^{J} \mathcal{N}(f_k|\mu_j, \Sigma_j)\pi_j$
4: **for** Each set F received **do**
5:     **for** $i = 1$ to J_{max} **do**
6:         Initialize μ_j, Σ_j, μ_j
7:         **while** Not converged **do**
8:             **E step**: evaluate: $\beta(Z_n)$; Eq. 5
9:             **M step** evaluate: μ_j, Σ_j, π_j ; Eqs. 6, 7, 8
10:           Check for convergence
11:         **end while**
12:     **end for**
13:     Find J_{opt} for minimum γ
14: **end for**

Thus, to design the LTKF, we use the center pixel velocities in the LTKF prediction model to estimate the landmarks image points, and we use the c_j measurements in the correction step. The relative velocities of the landmarks 3D points $^C\dot{\mathbf{X}}_j = [\dot{x} \ \dot{y} \ \dot{z}]^T$ are directly related to the NVS linear velocity $^L\mathbf{v} = [v_x \ v_y \ v_z]^T$, angular velocity $^L\boldsymbol{\omega} = [\omega_x \ \omega_y \ \omega_z]^T$, and the relative depth of each landmark Z_j defined in $\mathcal{F}_L$ by [39]

$$^C\dot{\mathbf{X}}_j = -^L\mathbf{v} - ^L\boldsymbol{\omega} \times ^C\mathbf{X}_j, \tag{11}$$

which expands to

$$\dot{x} = -v_x - \omega_y Z_j + \omega_z Y_j, \tag{12}$$
$$\dot{y} = -v_y - \omega_z X_j + \omega_x Z_j, \tag{13}$$
$$\dot{z} = -v_z - \omega_x Y_j + \omega_y X_j, \tag{14}$$

where X_j, Y_j, and Z_j are the landmarks relative positions defined in $\mathcal{F}_C$.

Accordingly, the landmarks pixel velocities $\dot{c}_j = [\dot{u} \ \dot{v}]^T$ are a function of their center pixel coordinates c_j, the relative depth of the landmarks Z_j, and the camera intrinsic matrix K as [39], [40]

$$\dot{c}_j = \mathbf{L}_{s,j}(u, v, Z_j, K) \begin{bmatrix} ^L\mathbf{v} \\ ^L\boldsymbol{\omega} \end{bmatrix}, \tag{15}$$

where $\mathbf{L_{s,j}} \in \mathbb{R}^{2 \times 6}$ is the interaction matrix for each landmark. Accordingly, taking the time derivative of $\dot{u}_i$, the interaction matrix for each flickering landmark $\mathbf{L_s}$ is given by [40]

$$\mathbf{L_s} = \begin{pmatrix} -\frac{f_u}{Z_j} & 0 & \frac{u}{Z_j} & \frac{uv}{f_v} & -f_u - \frac{u^2}{f_u} & \frac{f_u}{f_v}v \\ 0 & -\frac{f_v}{Z_j} & \frac{v}{Z_j} & f_v + \frac{v^2}{f_v} & -\frac{uv}{f_u} & -\frac{f_u}{f_v}u \end{pmatrix}.$$

Through the nonlinear Eq. 15 it is possible to fuse the measured c_j with the inertial states of the NVS. However, due to lower update rate of c_j, pixelation errors, and noises in the polarity transition events, such fusion is neglected and c_j is decoupled from NVS states in the LTKF. This would also allow us to have a computationally efficient filtration scheme

by using a much smaller state vector. Accordingly, we define the prediction model for the LTKF as

$$\tilde{\mathbf{c}}_{j,k+1} = \begin{bmatrix} \tilde{u}_{k+1} \\ \tilde{v}_{k+1} \end{bmatrix} \tag{16}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tilde{\mathbf{c}}_{j,k} + \begin{bmatrix} \Delta t \\ \Delta t \end{bmatrix} \mathbf{u}_{j,k} + \mathcal{N}(0, Q_p), \tag{17}$$

$$\text{where } j \in \{1, 2, ..., J_{opt}\},$$

where $\tilde{\mathbf{c}}_{j,k}$ is the predicted state vector comprising the landmarks image points, $\mathbf{u}_{j,k}$ is the control input vector of the estimated pixel velocities $\dot{\mathbf{c}}_j$, and Q_p is the covariance for an additive zero-mean Gaussian process noise. As customary for these types of filters, an update step is required for each of the LTKFs in $j \in \{1, 2, ..., J_{opt}\}$, where their measurement models are defined with the measurement vectors $\mathbf{z}_{j,k} = \mathbf{c}_j$ and the observation matrices $H_{j,k} = \mathcal{I}_{2 \times 2}$. It is important to highlight that $^L\boldsymbol{\omega}$ is obtained from the IMU gyroscope and $^L\mathbf{v}$ is estimated from the TDKF, which is discussed in the following section.

B. Translation Decoupled Kalman Filter

Given the tracked landmarks image points $\tilde{\mathbf{c}}_j$, it is possible to estimate the relative position of the NVS with respect to the landmarks by means of solutions to the PnP problem. Nevertheless, this does not provide the sufficient relative localization accuracy and the PnP positioning estimates need to be differentiated to obtain $^L\mathbf{v}$ which tends to be noisy, deteriorating the tracking performance of the LTKF. Therefore, the TDKF utilizes the IMU accelerometer measurements $^B\mathbf{a} = [a_x \ a_y \ a_z]$ in its prediction model and the PnP positioning estimates as measurement corrections for robustly estimating the $^L\mathbf{v}$ as well as the translation vector $^L\mathbf{t}$ pointing from $\mathcal{F}_C$ to $\mathcal{F}_L$.

It is important that the acceleration measurements are defined in $\mathcal{F}_L$ to be utilized in the TDKF. First, $^B\mathbf{a}$ is rotated to $\mathcal{F}_C$ with $^C_B\boldsymbol{R}$. The translational offset between frames $\mathcal{F}_C$ and $\mathcal{F}_B$ is neglected since the acceleration measurements are not affected by the NVS and IMU positions given that they are on the same rigid body. Then, $^C\mathbf{a}$ is transformed to $\mathcal{F}_L$ using the Madgwick orientation filter [41]. Nevertheless, it is substantial to provide a reliable source for the relative yaw ϕ of $\mathcal{F}_C$ with respect to $\mathcal{F}_L$ to account for yaw angle drift. Hence, ϕ is estimated by solving the PnP problem replacing the magnetometer in Madgwick's filter [41], demonstrated in Fig. 1.

After obtaining $^B\mathbf{a}$ in $\mathcal{F}_L$, the TDKF prediction model is defined as follows

$$\tilde{\mathbf{x}}_{k+1} = \begin{bmatrix} ^L\tilde{\mathbf{t}}_k \\ ^L\tilde{\mathbf{v}}_k \\ \tilde{\mathbf{a}}_{b,k} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \Delta t & -\Delta t^2 \\ 0 & 1 & -\Delta t \\ 0 & 0 & 1 \end{bmatrix} \tilde{\mathbf{x}}_k + \begin{bmatrix} \Delta t^2 \\ \Delta t \\ 0 \end{bmatrix} u_k + \mathcal{N}(0, \sigma_p), \tag{18}$$

where $\tilde{\mathbf{x}}_{k+1}$ is the TDKF predicted state vector comprising of the predicted NVS position $^L\tilde{\mathbf{t}}_k$ and velocity $^L\tilde{\mathbf{v}}_k$ along each axis, $a_{b,k}$ is an estimate for acceleration bias to adjust for gravity, and Q_p is the variance for an additive zero-mean Gaussian process noise.

The prediction model of the TDKF provides reliable relative pose estimates but measurement corrections are substantial to avoid drifts due to the integration of acceleration bias. Thus, solutions to the PnP problem providing are used to correct the predicted TDKF states.

Solutions to the PnP problem estimate the translation $^L\mathbf{t}_p$ and rotation $^L_C R_p$ of $\mathcal{F}_{NVS}$ with respect to $\mathcal{F}_L$ given $^L\mathbf{X}_j$ projected as $\mathbf{c}_j$ with the NVS perspective projection model as

$$s \begin{bmatrix} \mathbf{c}_j \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} fx & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{K} \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}}_{^L_C R_p | ^L\mathbf{t}_p} \begin{bmatrix} ^L\mathbf{X}_j \\ 1 \end{bmatrix},$$

where s is a scaling factor. Obviously, no algebraic solution exists for $^L_C R_p$ and $^L\mathbf{t}_p$, since for each projected point the number of unknown parameters exceeds the known terms. Hence, the Gauss Newton minimization scheme (EPnP-GN) [42] is utilized to estimate both parameters obtained by iteratively minimizing the reprojection error, where the states are parameterized with an exponential map and the perspective projection model transforms to

$$s \begin{bmatrix} \mathbf{c}_j \\ 1 \end{bmatrix} = K \exp(\zeta) \begin{bmatrix} ^L\mathbf{X}_j \\ 1 \end{bmatrix}. \tag{19}$$

From Eq. 19, a least-squares problem is formulated to minimize the reprojection error as a function of ζ:

$$\zeta = \arg\min_\zeta \frac{1}{2} \sum_i ||u_i - \frac{1}{s} K \exp(\zeta) X_w||^2 \tag{20}$$

After obtaining $^L\mathbf{t}_p$ and $^L_C R_p$, the TDKF measurement model is formulated with the measurement vector $\mathbf{z} = {}^L\mathbf{t}_p$ and observation matrix $H_k = [\mathcal{I}_{3 \times 3} \ 0_{3 \times 6}]$.

V. EXPERIMENTS

A. Experimental Setup

We applied our approach on a space application to perform relative localization between the Ingenuity and the Perseverance Rover. The Ingenuity shall be represented by an unmanned aerial vehicle (UAV) with onboard NVS and IMU, and the rover by an unmanned ground vehicle (UGV) with the flickering landmarks; the experimental architecture is demonstrated in Fig, 5. Accordingly, DJI F550 and Jackal were utilized as a UAV and a UGV, respectively, see Fig. 7. The experiments were performed on two scenarios: (1) UAV performing a take off and hover and (2) UAV following a square trajectory while observing the landmarks on the UGV. The experiments were carried out within a seven meter range due to the limited indoor space envelope. Videos of the experiments are available through the following link: https://youtu.be/N3Ypu95c3_4.
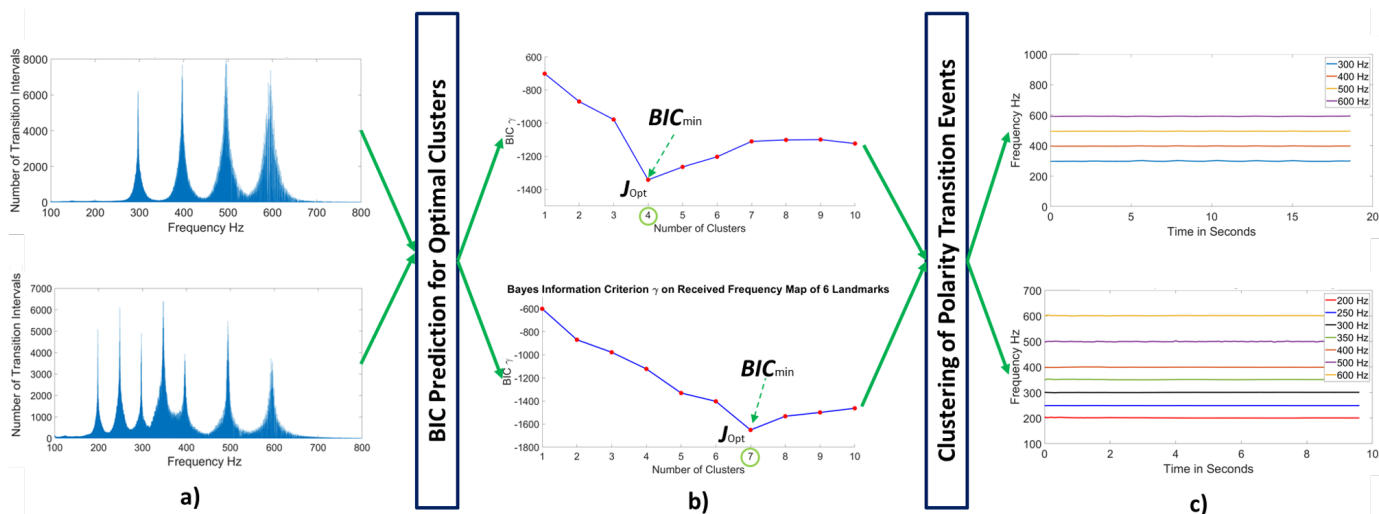
Fig. 4: a) The detected noisy polarity transition frequencies. b) The polarity transition events are clustered using GMMs and the BIC criterion extracts the optimal number of clusters corresponding to the observed number of landmarks. c) The clustered transition frequencies represent the true value of the landmarks flickering frequencies.
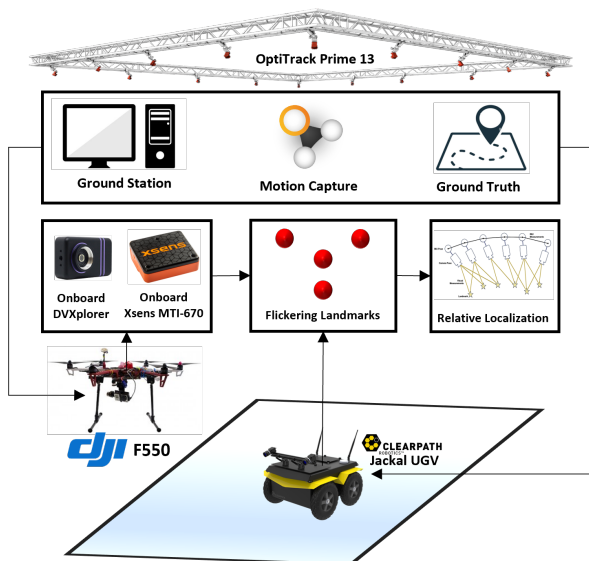

Fig. 5: Experimental protocol followed comprising of DJI F550 with onboard NVS and IMU and Jackal with flickering landmarks.

It is important to highlight that the UGV was assumed static since the Perseverance Rover speed (0.043 m/s) is relatively slow compared to the Ingenuity (10 m/s) [1]. This also satisfies the requirement on quasi static features given in section III. This assumption holds given the update rate of the LTKF measurement corrections (100 Hz), in which rising positioning error due to the assumption would be approximately 0.0004 m. Nevertheless, this can be easily surpassed with a constant velocity model adopted for the UGV. On the other hand, an Arduino Uno R3 ATmega16U2 was used to deliver unique frequency PWM signals to each landmark. Seven LedTech 228-5073 luminaries, of luminous intensity of 1300 mcd, were used as features flickering at 200, 250, 300, 350, 400, 500, and 600 Hz. The landmarks were placed at 1 meter apart

abiding by the Perseverance Rover dimensions (3 m × 3 m × 3 m). This is also substantial for robust relative localization and avoiding near planar constellations which introduce positioning flip ambiguities. The DVXplorer was utilized as the NVS and Xsens MTI-670 as the IMU, both interfaced with robot operating system (ROS) using a USB 3.0 terminal on the UAV onboard processor. Two important aspects that were taken into consideration during the experiments are the time and space asynchrony between the received measurement streams from the NVS and the IMU. The asynchronous measurements were synchronized using soft time synchronization by ROS [43] providing timestamped measurements from both sensors. The time offset between the two sensor clocks needed to be specified as a priori and was obtained through a step test with the times in which both sensors indicate that the UAV started flying have been compared. The time offset has been estimated to be 3.2 ms. On the other hand, the rotation between the NVS and IMU frames is substantial for the robustness of the system as discussed in section IV-B. This was obtained by converting NVS events to grayscale images using [44] and utilizing the approach in [45] to obtain the inertial measurements in the NVS reference frame. Finally, Optitrack Prime 13 was utilized to provide ground truth poses for both platforms.

B. Landmark Identification Results

Accurate frequency-based identification is critical to accurately formulate the NVBMs. The frequency detection error can be up to 33.97 Hz when relying on raw frequency measurements for landmark identification. This can lead the PnP solutions to diverge as the landmarks' correspondences are not matched correctly. On the other hand, we show that the unsupervised clustering of polarity transition events algorithm extracts the landmarks' true flickering frequencies and learns the optimal number of clusters for the polarity transition events. This was validated on the previously mentioned experimental setups and was also tested in two more scenarios where
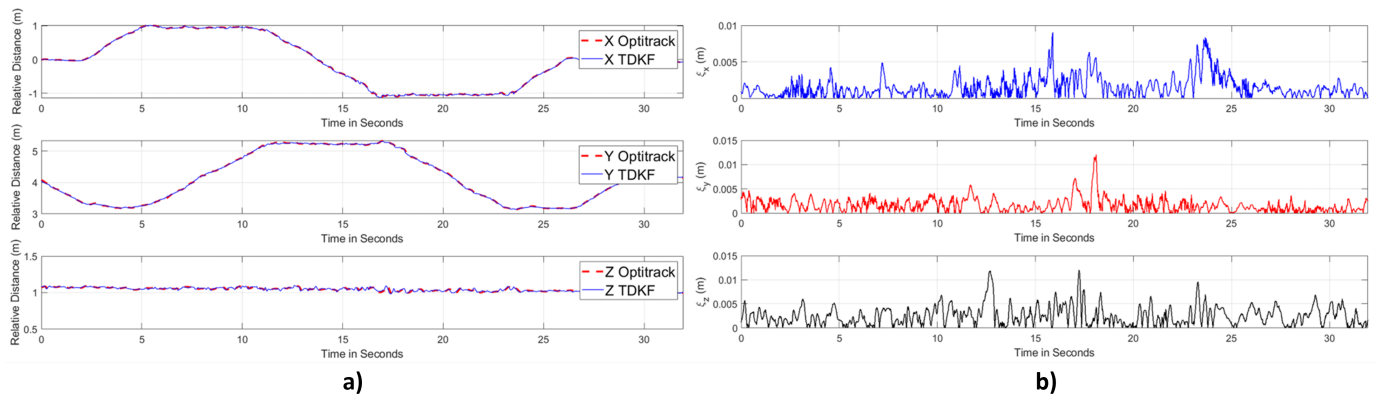
Fig. 6: a) UAV position relative to UGV estimated by TDKF against optitrack, where a maximum absolute positioning error $\xi_{x,y,z}^{max}$ of 0.0137 m and b) a maximum absolute relative orientation error $\xi_{\phi,\theta,\psi}^{max}$ of 2.16 degrees was achieved.
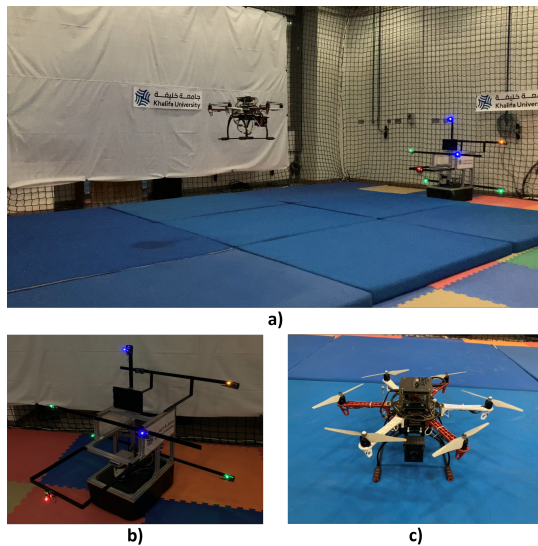

Fig. 7: a) Flying UAV observing flickering landmarks mounted on the UGV, b) Jackal with seven landmarks, and c) DJI F550 with DVXplorer and Xsens MTI-670 IMU.

a random hand-motion was induced on (1) four and (2) seven observed landmarks, illustrated in Fig. 4. When relying on our unsupervised clustering approach, the maximum frequency detection error dropped to 3.21 Hz. This is surely robust since the landmarks correspondences are matched correctly and the PnP problem solution is optimized. In addition, the minimum γ proved to be sufficient for finding J_{opt} corresponding to the number of observed landmarks.

C. Landmark Tracking Results

The LTKF provides a smoother and more robust landmark tracking especially in agile maneuvers compared to conventional imaging techniques relying on tracking by instance detection. We compare the LTKF to one of the most prominent feature extraction techniques for vision-based measurements, circle hough transform (CHT) [46], running at 100 Hz during take-off in the hovering scenario, see Fig. 8. We rely on linear interpolation for their comparison due to their heterogeneous

update rates. Note that the huge fluctuations are due to the non-optimal tuning of the UAV which caused it to vibrate.

The average difference between both trackers' outputs for a 600 Hz landmark along the Y axis in the image plane is 0.78 pixels which provides the required performance for robust relative localization between both platforms. It is still important to highlight that the LTKF is delayed compared to the CHT due to its inherent dynamics but does not exceed 0.0072 ms, which is negligible compared to its 5 ms sampling time.
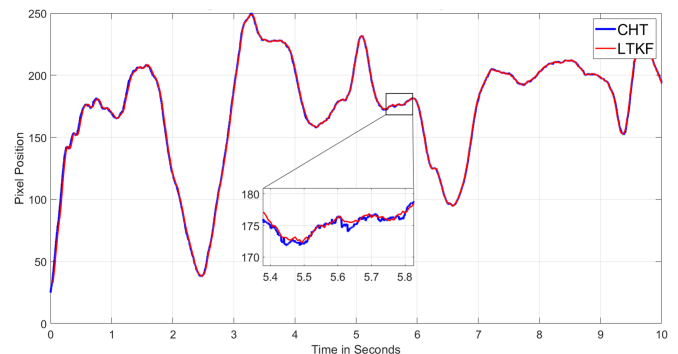

Fig. 8: Tracking plots for the CHT and LTKF on a 600 Hz landmark along X.

D. Relative Localization Results

We validate the accuracy of our TDKF against the ground truth from the optitrack motion capture system for the UAV's relative position to the UGV. We first show that relative positioning estimates of the CHT-PnP suffers from large positioning errors compared to the Optitrack and the LTKF-TDKF state estimators, illustrated in Fig. 9.

By only solving the PnP problem for relative localization, the positioning error can rise up to 0.09 m. This is due to the noise present in the CHT tracking performance and the relatively low resolution of the NVS. On the other hand, the LTKF-TDKF filters the noise providing more robust positioning estimates. Fig. 6 shows the TDKF estimated relative
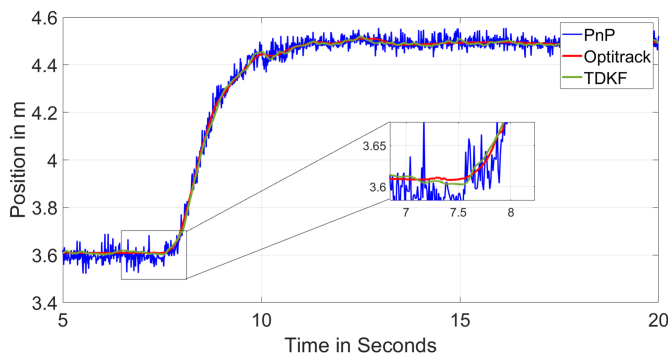
Fig. 9: Predicted relative position of PnP, TDKF, and optitrack demonstrating the noise present in CHT-PnP estimates and are filtered with the TDKF.

position compared to Optitrack for the scenario in which the UAV was following a square trajectory and Fig. 10 shows the trajectory followed relative to the UGV. Accordingly, a maximum relative positioning error $\xi_{x,y,z}^{max}$ of 0.0137 m and a mean relative positioning error $\bar{\xi}_{x,y,z}$ of 0.0052 m within 7 meter range were achieved with the TDKF. It is noticeable that the positioning error tends to increase at time instances where the UAV starts to move along an axis. This rises from the sudden induced motion (i.e. at $t = 17$ s in Y) but the TDKF converges quickly, yet the error remains within the acceptable range. On the other hand, the Madgwick filter fused with the yaw measurements obtained from the PnP solution was utilized for estimating the UAV orientation, achieving maximum orientation error $\xi_{\phi,\theta,\psi}^{max}$ of 2.16 degrees and a mean orientation error $\bar{\xi}_{\phi,\theta,\psi}$ of 0.567 degrees.
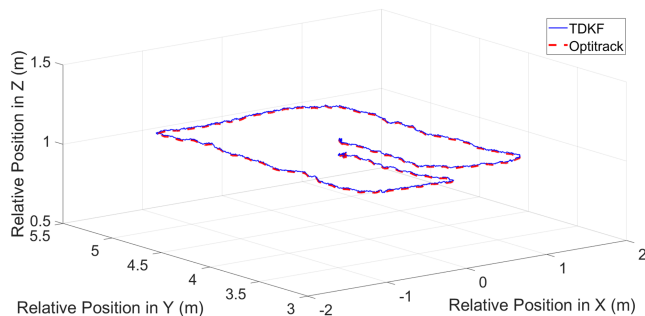

Fig. 10: Estimated UAV trajectory by the TDKF compared against optitrack.

The performance of the TDKF depends on the accuracy of the measurement corrections utilized from solving the PnP problem. If the landmarks were placed in a near planar constellation, multiple solutions to the PnP problem exist leading to undesired flip ambiguities, see Fig. 11. To avoid such ambiguities, we have conducted several experiments with various layouts of the landmarks' placement on the UGV. It was concluded that at least two landmarks need to be placed, such that the distance between them along the NVS optical axis is equal to one-seventh the maximum distance between the UAV and UGV. For our test scenario, the maximum distance between the UAV and UGV was 7 m and hence, two landmarks

needed to placed 1 m apart along the optical axis of the NVS, as shown in figure 7. Given the geometrical constraints of the Perseverance Rover, d_l can be at most 3 m. Hence, we conjecture that the relative localization range could increase to approximately 21 m. More importantly, landmarks with higher luminous intensities need to be utilized to ensure that they are detected from such a range. As 1300 mcd landmarks were detected from 7 meter range and since light propagation follows the inverse square law [47], the required illumination to trigger the NVS pixels must be 27 mcd. This implies that for landmarks to be detected from 20-meter range, the required illumination needs to be 10,600 mcd. It is fundamentally important to highlight that the measurement corrections for the LTKF are independent of the landmarks layout since they are clustered based on their unique transition frequencies regardless of their positions in the image plane as long as they are not completely occluded.
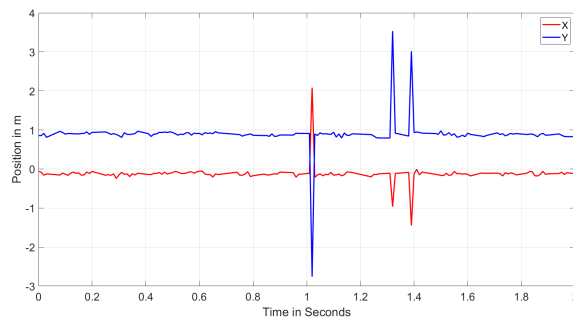

Fig. 11: Ambiguous solutions for the PnP problem rising from near planar landmarks layout resulting in large positioning errors.

E. Benchmarks

To evaluate the impact of our developed relative localization system, we compare our results with previous VBM-based relative localization systems from the literature. We benchmark our system performance against the works of Breitnmoser et al. [48], Faessler et al. [49], and Dias et al. [50] utilizing conventional imaging sensors to observe active LED luminaries for relative localization. We also compare our work against the work of Chen et al. [32] where an NVS is utilized with flickering markers for localization of a single moving platform. Table I lists our results compared to those works.

Our approach shows breakthrough performance for both relative position and orientation estimation. Compared to Faessler et al. [49], which provides the best localization accuracy out of all previous works with a higher resolution camera compared to ours, the developed system contributed to a 0.02 m and 1° improvement in positioning and orientation accuracy, respectively, within a longer range and a lower resolution sensor. We also benchmark our system against the approach developed by Chen et al. [32]. Even though we utilized a higher resolution NVS, the relative localization range was 7 times larger than theirs which made it more challenging to position the UAV. Hence, we show that the our sensor pixel coverage area is much larger compared to theirs. We assume

TABLE I: The developed relative localization system is compared to the works of Faessler et al., Chen et al. and Li et al. Our system surpasses the performance of previous SOTA works in estimating the relative pose of the platform within a larger range. The dash indicates that the metric was not mentioned in the article.

Authors	Breitenmoser et al. [48]	Faesslar et al. [49]	Dias et al. [50]	Chen et al. [32]	Ours
Year	2011	2014	2016	2020	**2022**
Vision Sensor	CMOS	CMOS	CMOS	NVS	**NVS**
Sensor Resolution	752 x 480	752 x 480	752 x 480	346 x 260	**640 x 480**
Maximum Positioning Error (m)	0.12	0.0328	0.25	0.03	**0.0137**
Mean Positioning Error (m)	0.015	0.0074	0.17	-	**0.0052**
Maximum Orientation Error (○)	4.5	3.37	12.5	-	**2.16**
Mean Orientation Error (○)	1.2	0.79	-	-	**0.567**
Range (m)	1.75	5.5	3.5	1	**7**
Field of View (○)	90	90	90	45	**45**
Pixel Coverage Area (mm^2)	213.16	213.2	26.0	5.76	**88.36**
Positioning Percentage Error	7%	0.13%	4.85%	3%	**0.074%**

that the FOV is $45°$ since the same lens is provided for both NVS models. The pixel coverage area is computed by finding the focal length f from the FOV

$$FOV = 2\tan^{-1}\left(\frac{W}{2f}\right), \tag{21}$$

where W is the width of the image plane and then the pixel area from the normalized image coordinates is computed given the range. Accordingly, the covered pixel area computed from the perspective projection model for our scenario tends to be 88.36 mm^2 much larger than in their approach, 5.76 mm^2.

On the other hand, the average system execution time is also reported for each system block, see Table II. The proposed approach shows real time applicability with a total execution time of 2.933 ms when compared to Faesslar et al. [49] whose total execution time is 3.8 ms. It is worth noting that the Faesslar et al. [49] approach was executed on C++ while ours was developed using Matlab on Intel i7-10750H (2.60GHz) processor. Thus, our execution time can be further improved if implemented using C++.

TABLE II: Execution times for the system blocks.

Algorithm	Execution Time (ms)
Polarity Transition Events Construction	0.003
Unsupervised Clustering of Polarity Transition Events	0.12
LTKF Prediction	0.47
Madgwick Filter	1.13
PnP Pose Estimation	0.97
TDKF Prediction	0.24
Total	**2.933**

VI. CONCLUSION

In this paper, a robust relative localization system based on NVBMs aided with flickering landmarks has been developed and was addressed with an application for the Ingenuity and Perseverance Rover. Flickering landmarks were adopted on the targeted agents to guarantee their continuous observability by the neuromorphic vision sensor onboard the aerial vehicle, even in absence of relative motion between both robots. This was attained by developing a novel unsupervised clustering approach for the events for landmark identification and by designing state estimators, LTKF and TDKF, for robust tracking and relative localization without relying on conventional imaging techniques. The system was validated in experiments resembling the targeted space scenario where the proposed approach outperformed state-of-the-art methods in terms of localization accuracy and execution time. Even though the developed relative localization system was deployed in a space application, its robust performance opens up future opportunities for deployment in a wide variety of applications, such as relative localization of robotic swarms in space, localization systems for indoor and underwater positioning systems due to the unavailability of GPS in such areas. Future work will extend the system to those applications and utilize motion models for observed agents in robotic swarms. In addition, the influence of the landmarks arrangement will be investigated for defining their optimal layout to further improve the system localization accuracy.
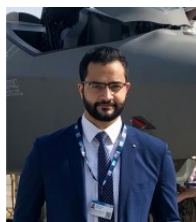
REFERENCES

[1] B. Balaram, T. Canham, C. Duncan, H. Grip, W. Johnson, J. Maki, A. Quon, R. Stern, and D. Zhu, "Mars helicopter technology demonstrator," 01 2018.

[2] S. Chen, D. Yin, and Y. Niu, "A survey of robot swarms relative localization method," *Sensors*, vol. 22, no. 12, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/12/4424

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2022.3217513

11

[3] S. Subedi and J.-Y. Pyun, "A survey of smartphone-based indoor positioning system using rf-based wireless technologies," *Sensors*, vol. 20, no. 24, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/24/7230

[4] F. Mahlknecht, D. Gehrig, J. Nash, F. M. Rockenbauer, B. Morrell, J. Delaune, and D. Scaramuzza, "Exploring event camera-based odometry for planetary robots," 2022.

[5] A. Motroni, A. Buffi, and P. Nepa, "A survey on indoor vehicle localization through rfid technology," *IEEE Access*, vol. 9, pp. 17 921–17 942, 2021.

[6] T. Sanpechuda and L.-o. Kovavisaruch, "A review of rfid localization: Applications and techniques," vol. 2, 06 2008, pp. 769 – 772.

[7] C. Li, E. Tanghe, P. Suanet, D. Plets, J. Hoebeke, E. De Poorter, and W. Joseph, "Reloc 2.0: Uhf-rfid relative localization for drone-based inventory management," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[8] Y. Zhang, X. Gong, K. Liu, and S. Zhang, "Localization and tracking of an indoor autonomous vehicle based on the phase difference of passive uhf rfid signals," *Sensors*, vol. 21, no. 9, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/9/3286

[9] G. Shi and Y. Ming, *Survey of Indoor Positioning Systems Based on Ultra-wideband (UWB) Technology*, 01 2016, pp. 1269–1278.

[10] A. Alarifi, A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M. A. Al-Ammar, and H. S. Al-Khalifa, "Ultra wideband indoor positioning technologies: Analysis and recent advances," *Sensors*, vol. 16, no. 5, 2016. [Online]. Available: https://www.mdpi.com/1424-8220/16/5/707

[11] J. Strader, Y. Gu, J. N. Gross, M. De Petrillo, and J. Hardy, "Co-operative relative localization for moving uavs with single link range measurements," in *2016 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 2016, pp. 336–343.

[12] S. Güler, M. Abdelkader, and J. S. Shamma, "Peer-to-peer relative localization of aerial robots with ultrawideband sensors," *IEEE Transactions on Control Systems Technology*, vol. 29, no. 5, pp. 1981–1996, 2021.

[13] Y. Alkendi, L. Seneviratne, and Y. Zweiri, "State of the art in vision-based localization techniques for autonomous navigation systems," *IEEE Access*, vol. 9, pp. 76 847–76 874, 2021.

[14] M. Yoshino, S. Haruyama, and M. Nakagawa, "High-accuracy positioning system using visible led lights and image sensor," 02 2008, pp. 439 – 442.

[15] B. Ghimire, J. Seitz, and C. Mutschler, "Indoor positioning using ofdm-based visible light communication system," in *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2018, pp. 1–8.

[16] W. Sakpere, M. Adeyeye Oshin, and N. Mlitwa, "A state-of-the-art survey of indoor positioning and navigation systems and technologies," *South African Computer Journal*, vol. 29, p. 145, 12 2017.

[17] G. Lee and H. Kim, "A hybrid marker-based indoor positioning system for pedestrian tracking in subway stations," *Applied Sciences*, vol. 10, no. 21, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/21/7421

[18] M. Cutler, B. Michini, and J. P. How, "Lightweight infrared sensing for relative navigation of quadrotors," in *2013 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2013, pp. 1156–1164.

[19] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 154–180, 2022.

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

[21] Y. Alkendi, R. Azzam, A. Ayyad, S. Javed, L. Seneviratne, and Y. Zweiri, "Neuromorphic camera denoising using graph neural network-driven transformers," 2021. [Online]. Available: https://arxiv.org/abs/2112.09685

[22] T.-J. Chin, S. Bagchi, A. Eriksson, and A. van Schaik, "Star tracking using an event camera," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 1646–1655.

[23] V. Kakani, X. Cui, M. Ma, and H. Kim, "Vision-based tactile sensor mechanism for the estimation of contact position and force distribution using deep learning," *Sensors*, vol. 21, no. 5, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/5/1920

[24] F. Baghaei Naeini, A. M. AlAli, R. Al-Husari, A. Rigi, M. K. Al-Sharman, D. Makris, and Y. Zweiri, "A novel dynamic-vision-based approach for tactile sensing applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 5, pp. 1881–1893, 2020.

[25] A. Rigi, F. Baghaei Naeini, D. Makris, and Y. Zweiri, "A novel event-based incipient slip detection using dynamic active-pixel vision sensor (davis)," *Sensors*, vol. 18, no. 2, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/2/333

[26] J. Bertrand, A. Yiğit, and S. Durand, "Embedded event-based visual odometry," in *2020 6th International Conference on Event-Based Control, Communication, and Signal Processing (EBCCSP)*, 2020, pp. 1–8.

[27] D. Liu, A. Parra, and T.-J. Chin, "Spatiotemporal registration for event-based visual odometry," 2021. [Online]. Available: https://arxiv.org/abs/2103.05955

[28] E. Mueggler, B. Huber, and D. Scaramuzza, "Event-based, 6-dof pose tracking for high-speed maneuvers," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 2761–2768.

[29] D. Reverter Valeiras, G. Orchard, S.-H. Ieng, and R. B. Benosman, "Neuromorphic event-based 3d pose estimation," *Frontiers in Neuroscience*, vol. 9, 2016. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2015.00522

[30] A. Nguyen, T.-T. Do, D. G. Caldwell, and N. G. Tsagarakis, "Real-time 6dof pose relocalization for event cameras with stacked spatial lstm networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 1638–1645.

[31] W. Yuan and S. Ramalingam, "Fast localization and tracking using event sensors," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 4564–4571.

[32] G. Chen, W. Chen, Q. Yang, Z. Xu, L. Yang, J. Conradt, and A. Knoll, "A novel visible light positioning system with event-based neuromorphic vision sensor," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10 211–10 219, 2020.

[33] A. Censi, J. Strubel, C. Brandli, T. Delbruck, and D. Scaramuzza, "Low-latency localization by active led markers tracking using a dynamic vision sensor," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 891–898.

[34] B. Li, H. Cao, Z. Qu, Y. Hu, Z. Wang, and Z. Liang, "Event-based robotic grasping detection with neuromorphic vision sensor and event-grasping dataset," *Frontiers in Neurorobotics*, vol. 14, p. 51, 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnbot.2020.00051

[35] X. Jianhong, G. Jianwen, W. Yan, L. Yuheng, L. Yongdong, C. Xiaochang, and Y. Guoliang, "Method of pose tracking by event camera using led marker points," in *2019 14th IEEE International Conference on Electronic Measurement Instruments (ICEMI)*, 2019, pp. 381–386.

[36] P. N. McMahon-Crabtree and D. G. Monet, "Commercial-off-the-shelf event-based cameras for space surveillance applications," *Appl. Opt.*, vol. 60, no. 25, pp. G144–G153, Sep 2021. [Online]. Available: http://opg.optica.org/ao/abstract.cfm?URI=ao-60-25-G144

[37] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, January 2006. [Online]. Available: https://www.microsoft.com/en-us/research/publication/pattern-recognition-machine-learning/

[38] K. Inthajak, C. Duanggate, B. Uyyanonvara, S. Makhanov, S. Barman, and T. Williamson, "Variable size blob detection with feature stability," 01 2010.

[39] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*. Berlin, Heidelberg: Springer-Verlag, 2007.

[40] I. S. Mohamed, G. Allibert, and P. Martinet, "Sampling-based MPC for constrained vision based control," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, sep 2021. [Online]. Available: https://doi.org/10.11092Firos51168.2021.9635970

[41] S. Madgwick, A. Harrison, and R. Vaidyanathan, "Estimation of imu and marg orientation using a gradient descent algorithm," *IEEE ... International Conference on Rehabilitation Robotics : [proceedings]*, vol. 2011, p. 5975346, 06 2011.

[42] L. Zhou and M. Kaess, "An efficient and accurate algorithm for the perspecitve-n-point problem," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 6245–6252.

[43] h. Hu, J. Wu, and Z. Xiong, "A soft time synchronization framework for multi-sensors in autonomous localization and navigation," in *2018 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, 2018, pp. 694–699.

[44] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "High speed and high dynamic range video with an event camera," 2019. [Online]. Available: https://arxiv.org/abs/1906.07165

[45] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 1280–1286.

[46] A. Shehata, S. Mohammad, M. Abdallah, and M. Ragab, "A survey on hough transform, theory, techniques and applications," 02 2015.

[47] N. Voudoukis, "Inverse square law for light and radiation: A unifying educational approach," *European Journal of Engineering Research and Science*, vol. 2, 11 2017.

[48] A. Breitenmoser, L. Kneip, and R. Siegwart, "A monocular vision-based system for 6d relative robot localization," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011, pp. 79–85.

[49] M. Faessler, E. Mueggler, K. Schwabe, and D. Scaramuzza, "A monocular pose estimation system based on infrared leds," 06 2014.

[50] D. Dias, R. Ventura, P. Lima, and A. Martinoli, "On-board vision-based 3d relative localization system for multiple quadrotors," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 1181–1187.

Abdulla Ayyad (Member, IEEE) received the M.Sc. degree in electrical engineering from The University of Tokyo, in 2019, where he conducted research with the Spacecraft Control and Robotics Laboratory. He is currently a Research Associate at the Advanced Research and Innovation Center (ARIC) at Khalifa University working on several robot autonomy projects. His current research interest includes the application of AI in the fields of perception, navigation, and control.

Mohammed Salah received his BSc. in Mechanical Engineering from American University of Sharjah, UAE, in 2020 and his MSc. in Mechanical Engineering from Khalifa University, Abu Dhabi, UAE, in 2022. He is currently with Khalifa University Center for Autonomous Robotic Systems (KUCARS). His research interest is mainly focused on multisensor fusion, neuromorphic vision, and space robotics.

Rana Azzam received the B.Sc. degree in computer engineering, the M.Sc. degree by Research in electrical and computer engineering, and the Ph.D. degree in engineering with a focus on robotics from Khalifa University in 2014, 2016, and 2020 respectively. She is currently a Postdoctoral Fellow with the Department of Aerospace Engineering. Her research interests include machine learning, reinforcement learning, navigation, and simultaneous localization and mapping.

Mohammed Chehadah (Member, IEEE) received his MSc. in Electrical Engineering from Khalifa University, Abu Dhabi, UAE, in 2017. He is currently with Khalifa University Center for Autonomous Robotic Systems (KUCARS). His research interest is mainly focused on identification, perception, and control of complex dynamical systems utilizing the recent advancements in the field of AI.

Lakmal Seneviratne is Professor of Mechanical Engineering and the founding Director of the Centre for Autonomous Robotic Systems (KUCARS) at Khalifa University, UAE. He has also served as Associate Provost for Research and Graduate Studies and Associate VP Research at Khalifa University. Prior to joining Khalifa University, he was Professor of Mechatronics, the founding Director of the Centre for Robotics Research and the Head of the Division of Engineering, at King's College London. He is Professor Emeritus at King's College London. His main research interests are centered on robotics and automation, with particular emphasis on increasing the autonomy of robotic systems interacting with complex dynamic environments. He has published over 400 peer reviewed publications on these topics. He is a member of the Mohammed Bin Rashid Academy of Scientists in the UAE.

Muhammad Humais received his M.Sc. in Electrical and Computer Engineering from Khalifa University in 2020. His research is mainly focused on robotic perception and control for autonomous systems. He is currently a Ph.D. fellow at Khalifa University Center for Autonomous Robotics (KU-CARS).

Mohammed Wahbah received his MSc. in Electrical Engineering from Khalifa University, Abu Dhabi, UAE, 2018. He is currently a researcher with Khalifa University Center for Autonomous Robotic Systems (KUCARS). His research areas include multisensor fusion, state estimation, and navigation in hazardous environments.

Yahya Zweiri (Member, IEEE) received the Ph.D. degree from the King's College London in 2003. He is currently an Associate Professor with the Department of Aerospace Engineering and deputy director of Advanced Research and Innovation Center - Khalifa University, United Arab Emirates. He was involved in defense and security research projects in the last 20 years at the Defense Science and Technology Laboratory, King's College London, and the King Abdullah II Design and Development Bureau, Jordan. He has published over 130 refereed journals and conference papers and filed ten patents in USA and U.K., in the unmanned systems field. His main expertise and research are in the area of robotic systems for extreme conditions with particular emphasis on applied Artificial Intelligence (AI) aspects and neuromorphic vision system.