Direct Georeferencing of Fire Front Aerial Images using Iterative Ray-Tracing and a Bearings-Range Extended Kalman Filter

Bernardo Santana, *MSc Student*, Instituto Superior Técnico

***Abstract*—This work proposes the Iterative Ray-Tracing as a forest fire georeferencing algorithm using images captured by an aerial vehicle. The Iterative Ray-Tracing requires that the vehicle is equipped with a Global Positioning System, an Inertial Measurement Unit and a Digital Elevation Map. In addition, the camera's Intrinsic Parameters and pixels to georeference must be known. Considering errors in the equipment, the Unscented Transform is proposed to characterize the uncertainty of the Iterative Ray-Tracing. Furthermore, a novel Bearings-Range filter measurement model is put forward, with the purpose of improving the target position and reducing its uncertainty. The georeferencing and filtering algorithms were validated with two simulations on rough and flat terrains. The standalone Iterative Ray-Tracing obtained Root Mean Squared Errors of 30.7 meters and 43.4 meters on the rough and flat terrains, respectively, while the Bearings-Range filtering algorithm achieves an improvement by reducing these errors to 11.0 meters and 19.9 meters, outperforming the Bearings-Only model. A performance comparison was done between the Extended and Cubature Kalman Filters using the Bearings-Range and Bearings-Only measurement models. Identical results were achieved, with the former performing slightly faster. Three experiments were conducted using real data and the results obtained demonstrate the applicability of the proposed methodology to georeference forest fires.**

***Index Terms*—forest fire, aerial vehicle, georeferencing, GPS, IMU, DEM**

I. Introduction

In recent years, remote sensing has seen an increased interest in the scientific community. The development and widespread of Unmanned Aerial Vehicles (UAVs) as a cheaper solution when compared to manned aerial vehicles enabled the development of target geolocalization applications. This includes precision agriculture [10] , natural disaster management [15] and fire detection and monitoring [9]. This work addresses the latter.

The most commonly used sensors regarding target geolocalization are digital cameras, Global Positioning Systems (GPSs) and Inertial Measurement Units (IMUs). This became known as Direct Georeferencing, since direct sensor orientation is computed by processing the information provided by the on-board sensors [8], i.e., the camera's extrinsic parameters (EPs) are directly calculated.

Depending on the accuracy requirements, sensors with different specifications can be used. However, the higher the desired accuracy, the higher the cost and/or size of the hardware. Due to payload constraints, UAV's typically use smaller and error-prone IMU's, specially in yaw measurements [4], such as MicroElectroMechanical Systems (MEMS). This lack of quality lead to the development of computer vision algorithms such as Structure from Motion (SfM) that extract and match features between images. The integration of these algorithms with the pose provided by the IMU and GPS greatly increase the accuracy of the georeferencing process.

Alternatively, Indirect Georeferencing can be used for systems that lack navigation equipment. This requires, however, the placement of Ground Control Points (GCP) to determine the camera's EPs, which can be time consuming and costly [7]. Furthermore, in a natural disaster scenario or in rough and inaccessible terrains, it is impracticable to place GCP.

I-A. Motivation

Fire propagation has been under investigation for many decades. The effect of environment variables such as terrain [12] has been studied and models for the fire propagation have been developed. However, the absence of a method capable of georeferencing the fire front, limits the usefulness of said models. Therefore, the aim of this work is to fill in this gap and develop an algorithm capable of performing forest fire geolocalization, enabling the use of fire propagation models to improve firefighting strategies. We also propose using the Unscented Transform (UT) to characterize the uncertainty of the algorithm, which can be very useful in making the stochastic simulation of fire propagation more realistic. Finally, a novel Bearings-Range measurement model is put forward, to improve the position and uncertainty estimates given multiple observations of the same target.

I-B. Related Work

Forlani et al. [5] apply direct georeferencing by using the on-board Global Navigation Satellite System with the Real-Time Kinematic option with SfM and Bundle Adjustment. No GCP are used. This methodology is, however, not suitable in a forest fire scenario, where

the lack of differentiated texture and smoke prevents feature extraction and matching.

Conte et al. [3] propose an image registration approach by pattern-matching the images collected from a Micro Aerial Vehicle with satellite imagery. Multiple measurements are taken and recursive least square filter is applied. Similarly to [5], this technique relies on feature extraction, and is therefore unreliable in a forest fire environment.

Ponda et al. [13] develop a Line-of-Sight Bearings-Only EKF for target localization. This requires, however, a prior knowledge of the target's position, which is not reviewed in that work. Xu et al. [16] propose the same measurement model using a CKF instead, considering possible linearization errors induced by the standard EKF. To determine an initial approximation of the target's position, the Iterative Photogrammetry (IPG) algorithm [14] is used. In spite of being efficient, the IPG method can diverge when the incidence angle is smaller than the profile inclination angle.

Leira et al. [11] propose the intersection of the optic ray with a flat surface. This generalization, however, is not suitable in rough terrains, as seen in [16].

II. Methodology

II-A. Camera Calibration

The camera calibration is an important procedure in computer vision that calculates the camera's intrinsic parameters and allows the extraction of metric information from bi-dimensional images. For this reason, before taking off, the camera must be calibrated, and assuming that the lenses configuration is not manipulated, these parameters remain constant throughout the flight.

Using one of the many computer vision tools available (OpenCV or MATLAB Image Processing and Computer Vision Toolbox), the calibration is done by moving and changing the pose of a known pattern in the camera's line of sight [17]. Usually this pattern is a checkerboard and the size of the checkerboard square is measured beforehand.

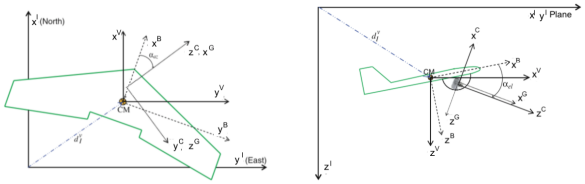
By performing the calibration, the camera's intrinsic (1) parameters are calculated, crucial to define the optic ray from the projection center to the target pixel.

$$K_{int} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where K_{int} is matrix of intrinsic parameters, $(f_x,\ f_y)$ is focal length and $(c_x,\ c_y)$ is the camera's principal point.

II-B. Coordinate Frames

Five coordinate frames were considered: camera, gimbal, body, vehicle and inertial. These frames are denoted respectively by: $F^C = (x^C, y^C, z^C)$, $F^G = (x^G, y^G, z^G)$, $F^B = (x^B, y^B, z^B)$, $F^V = (x^V, y^V, z^V)$ and $F^I = (x^I, y^I, z^I)$. The vehicle frame is a North-East-Down frame centered on the vehicle's centre of mass and is related to the inertial frame by a simple translation. If the coordinate transformations between all these frames are known by calibration, we can assume, without loss of generality, that the camera, gimbal and UAV centers of mass coincided with the GPS position, meaning that the translations in the frame transformations are zero.



(a) Top view (b) Side view

Fig. 1: UAV, gimbal and camera frames (adapted from [2])

II-B1 Camera Frame

The camera frame has its origin in the optical center, the x^C axis points to the right of the image plane , the y^C axis points downward on the image plane and z^C axis points in the direction of the optical axis of the camera. The rotation from the camera to the gimbal coordinate frame is defined by

$$R_C^G = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{2}$$

II-B2 Gimbal Frame

The gimbal coordinate frame has two degrees of freedom around y^G and z^G due to its pan and tilt movements, respectively. Defining the pan (elevation) and tilt (azimuth) angles as α_{el} and α_{az}, the rotation from the gimbal coordinate frame to body coordinate frame is given by $R_G^B = R_z(-\alpha_{az})R_y(-\alpha_{el})$,

$$R_G^B = \begin{bmatrix} c\alpha_{el}c\alpha_{az} & -s\alpha_{az} & s\alpha_{el}c\alpha_{az} \\ s\alpha_{az}c\alpha_{el} & c\alpha_{az} & s\alpha_{az}s\alpha_{el} \\ -s\alpha_{el} & 0 & c\alpha_{el} \end{bmatrix}, \tag{3}$$

where $c\gamma \triangleq \cos\gamma$ and $s\gamma \triangleq \sin\gamma$.

II-B3 Body Frame

The body frame describes the aircraft movement and has its origin in the center of mass of the UAV. The x^B axis points in the direction of the nose, the y^B axis points towards the right wing and the z^B axis points towards the aircraft belly. Defining the roll (ϕ), pitch (θ) and yaw (ψ) angles as the movement of the UAV around the axis x^B, y^B and z^B, respectively, the rotation from the body to the vehicle coordinate frame is defined as $R_B^V = R_z(-\psi)R_y(-\theta)R_x(-\phi)$,

$$R_B^V = \begin{bmatrix} c\psi c\theta & c\psi s\theta s\phi - s\psi c\phi & c\psi s\theta c\phi + s\psi s\phi \\ s\psi c\theta & s\psi s\theta s\phi + c\psi c\phi & s\psi s\theta c\phi - c\psi s\phi \\ -s\theta & c\theta s\phi & c\beta c\phi \end{bmatrix}, \tag{4}$$

.

II-C. Iterative Ray-Tracing

Defining $\vec{P}^V_{Target}$ as the vector pointing from the UAV to the target in the inertial coordinate frame and $\vec{P}^I_V$ as the vector pointing from the inertial frame origin to the UAV, the target position is defined as the sum of these vectors. To obtain $\vec{P}^V_{Target}$, the vector pointing from the camera optical center to the target pixel (u, v) must be defined, assuming $f_x \approx f_y$,

$$\vec{P}^C_{Target} = \left[u - c_x,\ v - c_y,\ \tfrac{f_x+f_y}{2}\right]^T, \quad (5)$$

and transformed to the vehicle coordinate frame using the coordinate frame transformations described in Section II-B,

$$\vec{P}^V_{Target} = [X'_d\ Y'_d\ Z'_d]^T = R^V_B R^B_G R^G_C \vec{P}^C_{Target}, \quad (6)$$

where $\vec{P}^V_{Target}$ references the target vector to the vehicle frame.

Given the camera position $\mathbf{R}_0 = [X_s\ Y_s\ Z_s]^T$ in the inertial frame and the normalized pointing vector

$$\mathbf{R}_d = [X_d\ Y_d\ Z_d]^T = \frac{1}{||\vec{P}^V_{Target}||} \cdot \vec{P}^V_{Target} \quad (7)$$

from the aerial vehicle to the target, in the vehicle frame, the ray $\mathbf{R}$ that starts in the vehicle and points to the target in the inertial frame is defined as

$$\mathbf{R}(t) = \mathbf{R}_0 + t \cdot \mathbf{R}_d = \begin{bmatrix} X_s \\ Y_s \\ Z_s \end{bmatrix} + t \cdot \begin{bmatrix} X_d \\ Y_d \\ Z_d \end{bmatrix}, \quad (8)$$

where t is the step and represents the distance between a point $\mathbf{R}(t)$ on the ray and the origin $\mathbf{R}_0$. When the ray elevation Z_{R} becomes less than the surface elevation Z_{DEM}, the intersection is detected. To do this, we use a method with a dynamic step size for t. The basic idea is the following. The algorithm is initialized with a large step value t and when the intersection is detected ($Z_{\mathrm{R}} \leq Z_{\mathrm{DEM}}$), the step size is reduced with a step divider, t_{div}, until the t becomes smaller than a pre-defined threshold, t_{th}. Furthermore, the starting iteration point, R'_0, is set as the intersection of the ray with the maximum elevation of the loaded DEM with respect to the inertial frame, Z_{max}. By defining a scaling factor,

$$\lambda = \frac{Z_{\max} - Z_{\mathbf{R}_0}}{Z_d}, \quad \text{for } \lambda > 0 \quad (9)$$

where $Z_{\mathbf{R}_0}$ is the height of the camera and Z_d is the third component of the normalized pointing vector, it is possible to calculate the new starting point, with Z_{R_0} = $Z_{\max}$

$$\mathbf{R}'_0 = \mathbf{R}_0 + \lambda \cdot \mathbf{R}_d. \quad (10)$$

Finally, bilinear interpolation was implemented, as in [16], to refine the elevation of the queried point. Ghandehari et al. [6] concluded in their work that for DEM's with finer resolutions, such has the EU-DEM v1.1 [1], the one used in this work, this type of interpolation achieves good results with low processing times.

Algorithm 1: Complete Iterative Ray-Tracing

Input : Telemetry from GPS and IMU, camera IP, target pixels, DEM
Output: Target coordinates in the inertial coordinate frame $\mathbf{t}_{\mathrm{I}}$ = $[x\ y\ z]$

```
Define ray origin R_0;
Define ray direction R_d;
Define step size t;
Define step size threshold t_th;
Define step size divider t_div;
if Z_{R_0} ≥ Z_max then
    Update ray origin with (10);
end
while No Intersection do
    Extend ray with (8);
    Interpolate ray;
    if Z_R ≤ Z_DEM AND t ≥ t_th then
        Update step t by diving it by step
         divider t_div;
    else if Z_R ≤ Z_DEM AND t < t_th then
        Intersection;
    else
        Augment t with current step value;
    end
end
```

II-D. Position and Uncertainty Characterization with the Unscented Transform

Three sources of uncertainty were taken into account: the vehicle GPS, IMU and gimbal. The GPS contributed with three degrees of uncertainty related to the position of the vehicle in the vehicle frame, σ_x, σ_y and σ_z. The vehicle IMU contributed with three degrees of uncertainty related to the orientation angles of the vehicle with respect to the vehicle frame, roll σ_ϕ, pitch σ_θ, and yaw σ_ψ. Finally, the gimbal contributed with two degrees of uncertainty related to the elevation and azimuth angles that establish the orientation of the gimbal with respect to the body frame, $\sigma_{\alpha_{el}}$ and $\sigma_{\alpha_{az}}$. This makes a total of $n = 8$ degrees of uncertainty. The standard deviations presented in Table I were assumed independent and the covariance matrix was defined as in (11).

$$\Sigma = \mathrm{diag}(\sigma_x^2, \sigma_y^2, \sigma_z^2, \sigma_\phi^2, \sigma_\theta^2, \sigma_\psi^2, \sigma_{\alpha_{el}}^2, \sigma_{\alpha_{az}}^2) \quad (11)$$

The UT parameters were set according to Table II. Since we approximate the distribution as a Gaussian, β = 2 is the optimal choice to minimize higher order information from the Taylor Series expansion. As for α and κ, these values were chosen so as to have the sigma points equal to the standard deviations of

Device	Standard Deviation σ_i	Value
GPS	σ_x	10m
	σ_y	10m
	σ_z	10m
IMU	σ_ϕ	1°
	σ_θ	1°
	σ_ψ	3°
Gimbal	$\sigma_{\alpha_{el}}$	1°
	$\sigma_{\alpha_{az}}$	1°

TABLE I: GPS, IMU and Gimbal standard deviations

Parameter	Value
α	$\frac{1}{\sqrt{8}}$
κ	0
β	2

TABLE II: UT Parameters

the equipment. The target position and uncertainty are calculated by propagating the sigma points with the complete iterative ray-tracing:

Algorithm 2: Unscented Transform with IRT

Output: Geodetic target coordinates $\mathbf{t} = [\phi\ \lambda\ h]$ and uncertainty $\sigma_{xyz} = [\sigma_x\ \sigma_y\ \sigma_z]$

Generate $2n+1$ sigma points;
for $i \leftarrow 0$ **to** $2n$ **do**
  Calculate the intersection for each sigma point, $y_i = \mathrm{IRT}(\chi_i)$;
end
Recover mean $\mathbf{t}$ and covariance $\mathbf{P}$;
Convert mean to geodetic coordinates and uncertainty $\sigma_{xyz} = \sqrt{\mathrm{diag}(\mathbf{P})}$;

II-E. Bearings-Range Measurement Model

Considering measurement errors in the GPS and IMU, this section details the proposed vision-based target localization using bearings and range measurements. Regarding previous works [13, 16], a new filter measurement model was developed that takes advantage of the available range information between the vehicle and the target. Taking into account the possible linearization errors induced by the EKF, a performance comparison is done with a CKF. Details on these filters can be found in [13, 16].

The bearings localization problem is based on the extraction of the azimuth β and elevation ϕ angles from the pointing vector, as shown in Figure 2. Since the proposed georeferencing algorithm calculates the 3D coordinates of the target, it enables the estimation of the distance r between the vehicle and the target.

A discrete non-linear system with Gaussian noise can be characterized as follows:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k) + \mathbf{w}_k \tag{12a}$$

$$\mathbf{z}_{k+1} = \mathbf{h}(\mathbf{x}_{k+1}) + \mathbf{v}_{k+1} \tag{12b}$$

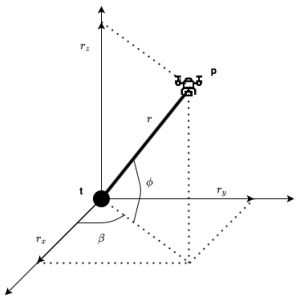


Fig. 2: Azimuth (β), elevation (ϕ) and range (r) between vehicle ($\mathbf{p}$) and target ($\mathbf{t}$)

The measurement model described in (12b) is given by:

$$\mathbf{h}(\mathbf{x}_k) = \begin{bmatrix}\beta\\ \phi\\ r\end{bmatrix} = \begin{bmatrix}\tan^{-1}\left(\frac{r_y}{r_x}\right)\\ \tan^{-1}\left(\frac{r_z}{\sqrt{r_x^2+r_y^2}}\right)\\ \sqrt{r_x^2+r_y^2+r_z^2}\end{bmatrix} \tag{13}$$

where $\mathbf{r}_k = [r_x\ r_y\ r_z]_k^T = [p_x - t_x,\ p_y - t_y,\ p_z - t_z]_k^T$ is the relative vector between the vehicle and the target for the k^{th} measurement, $\mathbf{p}_k = [p_x\ p_y\ p_z]_k^T$ is the position of the vehicle and $\mathbf{x}_k = \mathbf{t}_k = [t_x\ t_y\ t_z]_k^T$ is the position of the target and the state to be estimated.

Assuming that the fire front is stationary when comparing images taken within a small time interval, the target dynamics model described in (12a) is given by:

$$\mathbf{f}(\mathbf{x}_k) = \Phi_{k+1|k}\mathbf{x}_k = \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\mathbf{x}_k \tag{14}$$

and the covariance of the system noise w_k is given by

$$\mathbf{Q}_k = \begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix}. \tag{15}$$

For the EKF, the Jacobian of the measurement function with respect to the state is described in (16),

$$\mathbf{H}_k = \begin{bmatrix}\frac{r_y}{r_x^2+r_y^2} & -\frac{r_x}{r_x^2+r_y^2} & 0\\ \frac{r_x r_z}{||r||^2\sqrt{r_x^2+r_y^2}} & \frac{r_y r_z}{||r||^2\sqrt{r_x^2+r_y^2}} & -\frac{\sqrt{r_x^2+r_y^2}}{||r||^2}\\ -\frac{r_x}{||r||} & -\frac{r_y}{||r||} & -\frac{r_z}{||r||}\end{bmatrix}_k. \tag{16}$$

The sensor noise mentioned in Section II-D is used to model the noise covariance matrix, tuned to the following values

$$\mathbf{R}_k = \begin{bmatrix}\sigma_{\alpha_{az}}^2 & 0 & 0\\ 0 & \sigma_{\alpha_{el}}^2 & 0\\ 0 & 0 & \sigma_r^2\end{bmatrix} = \begin{bmatrix}1^2 & 0 & 0\\ 0 & 1^2 & 0\\ 0 & 0 & 10^2\end{bmatrix}, \tag{17}$$

where $\sigma_{\alpha_{az}}[^\circ]$ and $\sigma_{\alpha_{el}}[^\circ]$ are the gimbal's azimuth and elevation uncertainties and $\sigma_r[\mathrm{m}]$ is the range uncertainty. The IRT and UT results of the first observation initialize the filters' state and covariance, $\hat{\mathbf{x}}_\mathbf{0}$ and $\hat{\mathbf{P}}_0$.

II-F. Metrics

The metrics presented in this Section will be used throughout the following chapters. The position error, e_p defined in (18) is used to determine the distance between the target position, t, and its estimates, $\hat{t}$,

$$e_p = [t_x \ t_y \ t_z]^T - [\hat{t}_x \ \hat{t}_y \ \hat{t}_z]^T. \tag{18}$$

To characterize the accuracy of the algorithm, the average position error and RMSE are defined, respectively,

$$\mu_{e_p} = \frac{\sum_{i=1}^{N} \sqrt{e_{p_{x_i}}^2 + e_{p_{y_i}}^2 + e_{p_{z_i}}^2}}{N}, \tag{19}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} e_{p_{x_i}}^2 + e_{p_{y_i}}^2 + e_{p_{z_i}}^2}{N}}. \tag{20}$$

While the former provides the average distance between the estimates and the target, the latter characterizes their distribution.

Finally, the uncertainty is defined as the square root of the diagonal of the matrix $\mathbf{P}$,

$$\mathbf{P} = \begin{bmatrix} \sigma_{xx}^2 & \sigma_{xy}^2 & \sigma_{xz}^2 \\ \sigma_{yx}^2 & \sigma_{yy}^2 & \sigma_{yx}^2 \\ \sigma_{zx}^2 & \sigma_{zy}^2 & \sigma_{zz}^2 \end{bmatrix}, \tag{21}$$

$$\sigma_{xyz} = [\sigma_{xx}, \ \sigma_{yy}, \ \sigma_{zz}]^T. \tag{22}$$

The average position uncertainty is defined as

$$\mu_{\sigma_{xyz}} = \frac{\sum_{i=1}^{N} \sqrt{\sigma_{xx_i}^2 + \sigma_{yy_i}^2 + \sigma_{zz_i}^2}}{N}. \tag{23}$$

III. Simulation Procedure

In this work, two terrain types were considered, rough and flat, as in previous works on this topic [16]. In both cases, a linear trajectory was followed, with the aerial vehicle flying at a speed of 250 km/h. Ground truth telemetry was generated and perturbed with the following model:

$$\hat{\mathbf{x}}_i(t) = \mathbf{x}_i(t) + \mathbf{w}(t), \quad \mathbf{w} \sim \mathcal{N}(0, \sigma_{x_i}), \tag{24}$$

where $\hat{\mathbf{x}}_i$ is the noisy variable, $\mathbf{x}_i$ is the ground truth variable, $\mathbf{w}$ is the zero-mean noise with σ_{x_i} standard deviation. The assumed values for σ_{x_i} are the same as the ones used for the UT, presented in Table I. For each simulation, 100 runs were performed with independent noise sequences, and the EKF and CKF performances using the Bearings-Range (BR-EKF, BR-CKF) and Bearings-Only (BO-EKF, BO-CKF), the latter described in detail in [13], measurement models were compared. The results presented are an average of all the runs.

III-A. Rough Terrain Simulation

A total of 25 measurements were collected at a constant height of 1650 meters. Figure 3 shows the resulting measurements following the model presented in (24) to disturb the system. The average distance to the target was 880 meters.

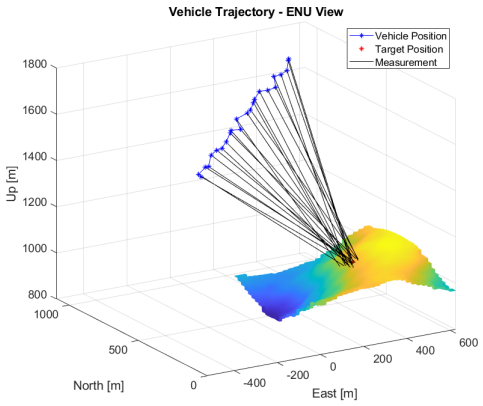


Fig. 3: Measurements with noise-induced position and attitude on the rough terrain simulation

Method	μ_{e_p} [m]	RMSE [m]	$\mu_{\sigma_{xyz}}$ [m]
IRT	30.527	30.743	35.965
IRT+BR-EKF	**6.221**	**11.726**	3.837
IRT+BO-EKF	8.225	13.071	5.137
IRT+BR-CKF	6.222	11.785	**3.782**
IRT+BO-CKF	8.193	13.136	5.054

TABLE III: IRT, IRT+BR-EKF, IRT+BO-EKF, IRT+BR-CKF and IRT+BO-CKF results for the rough terrain scenario

III-B. Flat Terrain Simulation

A total of 21 measurements were collected at a constant height of 950 meters. Figure 5 shows the resulting measurements with noisy data. The average distance to the target was 985 meters.

Method	μ_{e_p} [m]	RMSE [m]	$\mu_{\sigma_{xyz}}$ [m]
IRT	43.121	43.405	48.373
IRT+BR-EKF	11.020	19.910	**3.898**
IRT+BO-EKF	11.565	21.578	3.909
IRT+BR-CKF	**11.000**	**19.842**	3.899
IRT+BO-CKF	11.575	21.573	3.911

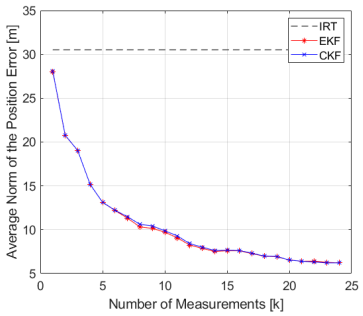
TABLE IV: IRT, IRT+BR-EKF, IRT+BO-EKF, IRT+BR-CKF and IRT+BO-CKF results for the flat terrain scenario

Kalman Filter	Average Processing Time [ms]
Extended	**0.795**
Cubature	0.83

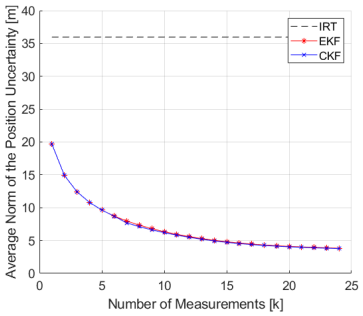
TABLE V: EKF and CKF processing time comparison

III-C. Discussion of Simulation Results

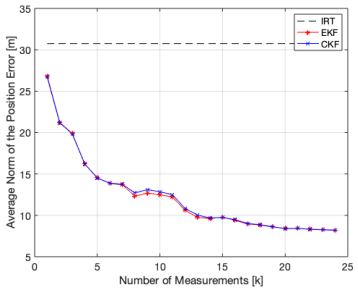
These simulations demonstrate the advantage of including the range information in the filtering algorithm. Both in the rough and flat simulations, the Bearings-Range measurement model achieves lower position errors, position uncertainties and RMSE's for the same number of measurements, therefore is more accurate than the Bearings-Only measurement model. Furthermore, it has a faster convergence, evident for $k = 10$ in Figures 4 and 6. The results presented in Tables III and IV show the clear improvement on the accuracy of the estimated target position when applying the Bearings-



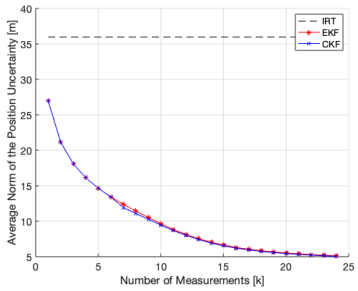
(a) BR-EKF and BR-CKF average norm of the position error



(b) BR-EKF and BR-CKF average norm of the position uncertainty



(c) BO-EKF and BO-CKF average norm of the position error



(d) BO-EKF and BO-CKF average norm of the position uncertainty

Fig. 4: EKF and CKF filtering results for the rough terrain simulation

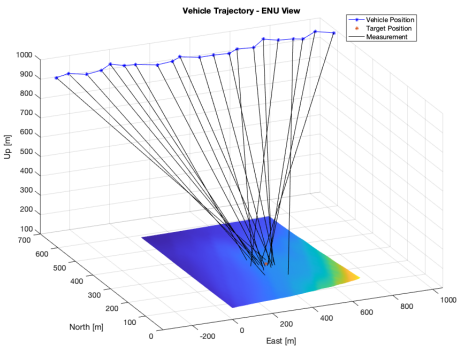


Fig. 5: Measurements with noise-induced position and attitude on the flat terrain simulation

Range filtering algorithm. The IRT RMSE is reduced by 61.86% and 54.12% for the rough and flat terrains, respectively, both for the BR-EKF and BR-CKF. Since both achieve the same final result with an identical progression in the position error and uncertainty, there is no clear advantage in using the BR-CKF over the BR-EKF. In addition, the EKF shows a slightly faster processing time per iteration (Table V), making it more appropriate for real-time applications.

IV. Experimental Procedure

Three experiments compose this section and are presented in the following manner: Subsection IV-A details the procedure where a mobile phone was used to acquire telemetry and imagery of a target along a pedestrian path, near *Porto de Mós*, *Leiria*.

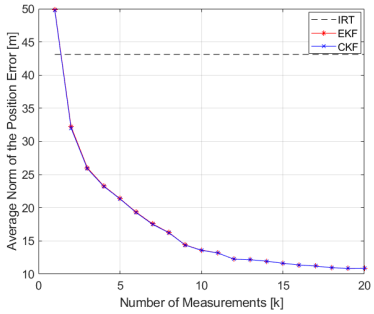
Subsection IV-B describes the georeferencing algorithm applied to a footage recorded by a Portuguese Air Force UAV near *Chaves*, *Vila Real*.

Finally, subsection IV-C describes another experiment using footage from a UAVision's UAV recorded near *Pombal*, *Leiria*. A minimization problem is proposed to calculate the camera's intrinsic parameters and the results are used on the georeferencing algorithm.

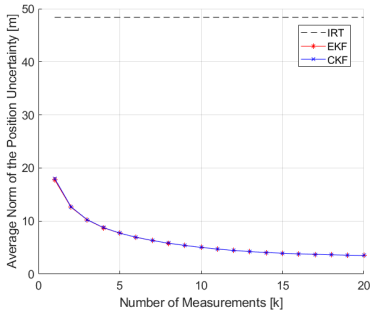
IV-A. Mobile Phone Procedure

A mobile phone was used to record GPS, IMU and image data along a pedestrian path. The natural elevation of *Serra dos Candeeiros*, near *Porto de Mós*, *Leiria*, was used to capture images of a target at a lower height, to simulate the overview perspective of an aerial vehicle. A total of 14 images were acquired at an average target distance of 605 meters. For the filtering, the IRT result for the first observation is used to initialize the filter state. For this experiment, considering that the IMU of the mobile phone had a lower quality when compared to the ones used onboard of an aerial vehicle, the measurement noise covariance matrix $\mathbf{R}_k$ was tuned to

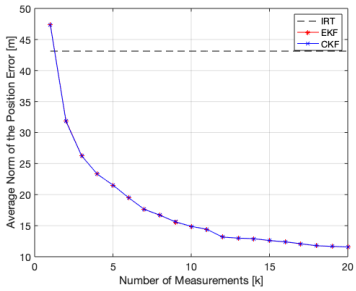
$$\mathbf{R}_k = \begin{bmatrix} 5^2 & 0 & 0 \\ 0 & 5^2 & 0 \\ 0 & 0 & 10^2 \end{bmatrix}. \tag{25}$$



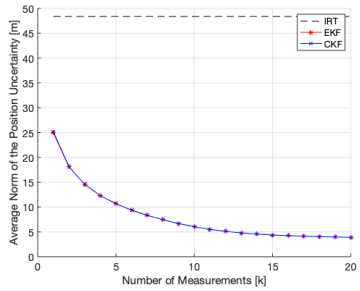
(a) BR-EKF and BR-CKF average norm of the position error



(b) BR-EKF and BR-CKF average norm of the position uncertainty



(c) BO-EKF and BO-CKF average norm of the position error



(d) BO-EKF and BO-CKF average norm of the position uncertainty

Fig. 6: EKF and CKF filtering results for the flat terrain simulation

IV-B. Portuguese Air Force UAV Footage

In this video, a UAV recorded a fire burning near *Chaves*, *Vila Real*, at 41.631724N -7.465919E. The camera intrinsic parameters were not available. However, the target was locked in the center of the image, making it possible to geolocate without them. The gimbal's azimuth and elevation angles are overlayed on the video feed. The former is calculated with respect to the vehicle coordinate frame. The latter, however, is provided with respect to the UAV's body. This was not problematic since the horizon line was parallel to the horizontal image frame. The feed also provided the UAV's position in the WGS84 reference system. A total of 7 frames were collected and used to test the georeferencing algorithm.

IV-C. UAVision UAV Footage

This video recorded a forest fire near *Pombal*, *Leiria*, at 39.832856N -8.519885E. As in the previous recording, the overlay displayed information on the azimuth and elevation angles, as well as the UAV's position in the WGS84 reference frame.

Again, the camera was not calibrated, therefore no intrinsic parameters were available. Furthermore, the vehicle performs a slight turn along its trajectory, meaning that the horizon line is not centered and parallel to the horizontal image frame and therefore the elevation angle is not reliable. To overcome these setbacks, three video frames with visible landmarks with known coordinates were selected. This allowed the formulation of a minimization problem to estimate the intrinsic parameters and to refine the rotation matrix of the system when those images were captured. As a cost function, the reprojection error was used

$$J(R_i, t_i, K_i nt) = \sum_{i=1}^{3} \sum_{k=1}^{N_k} (u_{ik} - \hat{u}_{ik})^2 + (v_{ik} - \hat{v}_{ik})^2, \quad (26)$$

where R_i and t_i are the rotation matrix and translation vector that establish the transformation from world to camera coordinates for frame i, K_{int} is the IP matrix, (u_{ik}, v_{ik}) is the measured pixel k in frame i and $(\hat{u}_{ik}, \hat{v}_{ik})$ is the predicted pixel using the current estimates of R_i, t_i and $K_i nt$. It was assumed that the skew s was zero, $f_x = f_y$ and $c_x = 640$ and $c_y = 360$:

$$\begin{aligned} \arg\min \quad & J(R_i, t_i, K_{\text{int}}) \\ \text{s.t.} \quad & ||r_{i1}||{=}1, \quad ||r_{i2}||{=}1, \quad ||r_{i3}||{=}1, \\ & r_{i1} r_{i2}^T{=}0, \quad r_{i2} r_{i3}^T{=}0, \quad r_{i1} r_{i3}^T{=}0, \\ & f_x{=}f_y, \end{aligned} \quad (27)$$

where r_{ik} is the row k of rotation matrix i. The six initial constraints are related to the orthogonality condition of the rotation matrices. The resulting intrinsic parameters and refined orientation were used to calculate the location of some landmarks.

V. Discussion of Results

V-A. Mobile Phone Experimental Results

A total of 20 images were used to calibrate the camera. As a result, the following intrinsic and distortion

parameters were obtained:

$$K_{\text{int}} = \begin{bmatrix} 3363.507 & 0 & 1967.377 \\ 0 & 3369.501 & 1419.890 \\ 0 & 0 & 1 \end{bmatrix} \quad (28)$$

$$\mathbf{k} = \begin{bmatrix} 0.2265 & -1.0227 & 1.7296 \end{bmatrix}^T \quad (29)$$

$$\mathbf{p} = \begin{bmatrix} -0.0098 & -0.0065 \end{bmatrix}^T \quad (30)$$

A mean reprojection error of 0.68 pixels was achieved with a 3-parameter radial distortion model.

The average distance between the position of the images and target was 640.83 meters, with a maximum of 770.92 meters and 493.90 meters. The results are presented in Table VI.

Method	μ_{e_p} [m]	RMSE [m]	$\mu_{\sigma_{xyz}}$ [m]
IRT	74.483	77.498	157.035
IRT+BR-EKF	**33.620**	**41.501**	**7.250**

TABLE VI: Standalone IRT and IRT+BR-EKF filtering results of the mobile phone experimental procedure

Figure 7 provides a clear view of the IRT estimated positions and the improvement achieved by applying the BR-EKF. A bias along the positive East direction is visible, since 12 of the 14 estimated positions are on that area. This can be due to the non-ideal experimental

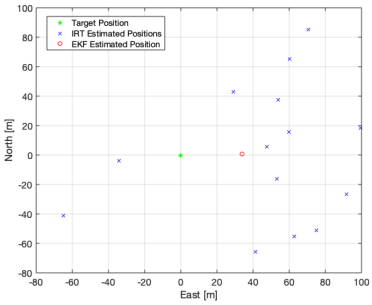


Fig. 7: Real and estimated target positions by the IRT and EKF algorithms

setup, i.e., a line-of-sight more parallel to the ground when compared to a more vertical one from an aerial vehicle. The mean position and uncertainty are plotted in Figure 8 and evidence this ill-conditioned setup, with a stretched uncertainty region along the line-of-sight direction. This means that a small error on the vertical image direction, be it the selected pixel or the roll angle, is amplified by this configuration. In addition, the images were taken at approximate positions, limiting the new information added to the BR-EKF.

The same experiment was simulated to investigate the origin of this bias. More specifically, to understand if it was due to the non-ideal perspective or if it was related to IMU or pixel identification errors. The mean position and uncertainty are plotted in Figure 9.

The results presented in Table VII demonstrate the

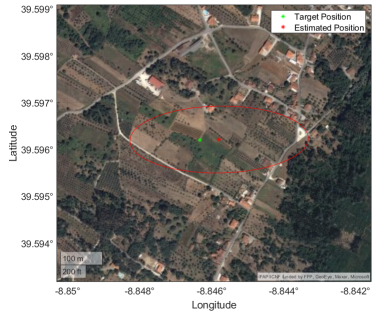


Fig. 8: Mean position and uncertainty from experimental procedure

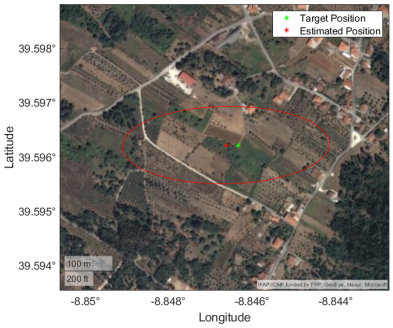


Fig. 9: Mean position and uncertainty from ground perspective simulation

Method	μ_{e_p} [m]	RMSE [m]	$\mu_{\sigma_{xyz}}$ [m]
IRT	134.377	140.2487	229.571
IRT+BR-EKF	**73.923**	**94.686**	**4.055**

TABLE VII: Results from the Porto de Mós experimental procedure simulated with a ground perspective

non-ideal perspective of the experimental procedure, with an increase by almost 100% in the RMSE and 72 meters in the position uncertainty when compared to Table VI. Furthermore, the mean estimated target position presented Figure 9 is shifted towards West when compared to the actual target position, showing that the bias identified in the experimental procedure, despite being towards East, is mainly caused by the sub-optimal line-of-sight.

Finally, the same trajectory was simulated with an aerial perspective at a height of 950 meters. The results presented in Table VIII validate the use of aerial vehicles when applying the proposed methodology. The RMSE decreased by 36 meters and the position

Method	μ_{e_p} [m]	RMSE [m]	$\mu_{\sigma_{xyz}}$ [m]
IRT	47.120	47.605	43.560
IRT+BR-EKF	**13.908**	**21.946**	**5.109**

TABLE VIII: Results from the Porto de Mós experimental procedure simulated with an aerial perspective

uncertainty was reduced by 113 meters, as a result of a more perpendicular line-of-sight evidenced in Figure 10. In addition, the biases identified in the experimental procedure and ground simulation are no longer present

since the mean estimated target position is practically identical to the real target position. In conclusion, the use of an aerial vehicle is clearly advantageous when applying the proposed algorithm, providing a more accurate estimate of the target's position, which is further improved with the Bearings-Range EKF.

Fig. 10: Mean position and uncertainty from aerial perspective simulation

V-B. Portuguese Air Force UAV Footage Results

In this experiment, the UAV flew at approximately 1920 meters, and the average distance to the target was 3183 meters, meaning that the horizontal component was 2539 meters. As in the *Porto de Mós* experimental procedure, the optic ray is more parallel to the ground when compared to the perspective of an aerial vehicle.

The fact that the observations are very similar means that no new information is added on each EKF iteration. Furthermore, only 7 measurements were taken, where one was used to initialize the filter state and covariance and only six were left to iterate. These factors contribute to the increase in the position error estimated by the filtering algorithm, as presented in Table IX. An important conclusion can be drawn from this. When georeferencing targets at greater distances, the uncertainties associated with the estimated positions are amplified. Therefore, the number and variability of measurements required for an accurate target position increases.

Method	μ_{e_p} [m]	RMSE [m]	$\mu_{\sigma_{xyz}}$ [m]
IRT	**136.250**	**146.049**	292.600
IRT+BR-EKF	198.434	157.056	**24.7578**

TABLE IX: Portuguese Air Force IRT and IRT+BR-EKF results

V-C. UAVision UAV Footage Results

The estimated intrinsic parameters from the minimization problem were

$$K_{\text{int}} = \begin{bmatrix} 1063.17 & 0 & 640 \\ 0 & 1063.17 & 360 \\ 0 & 0 & 1 \end{bmatrix}. \quad (31)$$

These parameters and the refined rotation matrices were used to georeference the position of several landmarks. Table X presents the number of landmarks selected in each frame and the average norm of the position error, average norm of position uncertainty and RMSE. Frames 1 and 2 present similar results in terms of position error and RMSE. Position uncertainty is greater on the first frame because it is georeferencing targets that are further from the vehicle than in the second frame. The third and final frame presents the worst results in terms of position error and RMSE. This is due to being the image with less landmarks, and as a consequence is less refined by the minimization problem when compared to the other frames. Overall, accurate results were obtained for targets that distanced more than one kilometer from the UAV. In a real scenario, were the position error and RMSE will not be available, the uncertainty is taken into account as metric of the georeferencing algorithm, with a lower value representing a more trustworthy estimated position.

Frame	Landmarks	μ_{e_p} [m]	RMSE [m]	$\mu_{\sigma_{xyz}}$ [m]
1	8	15.166	16.432	81.277
2	6	6.598	6.857	60.226
3	4	25.908	26.029	65.896

TABLE X: UAVision footage results

VI. Conclusions

The IRT algorithm proposed in this work presents a robust solution for forest fire georeferencing. It only requires the camera's IPs, the onboard sensor data (GPS/IMU/gimbal) and a DEM, while current state-of-the-art approaches rely on feature identification and matching.

Equipped with the UT, the georeferencing algorithm provides an estimate of the target position and characterizes its uncertainty. That is a relevant achievement since it provides crucial information regarding the confidence of the estimated target position to the firefighting personnel.

Finally, a novel Bearings-Range filter measurement model was developed to filter the bearings and range information from the relative geometry between the aerial vehicle and the estimated position.

The results obtained in a controlled simulation environment demonstrate the potential of the developed algorithm. At distances of almost 1000 meters, two simulations were run on a rough and flat terrains, achieving RMSEs of 30.7 and 43.4 meters with 25 and 21 measurements, respectively. After applying the BR-EKF, these errors reduced to 11.0 and 19.9 meters, a decrease of 58.5% and 54.7%. This relation was also verified in the mobile phone experimental procedure, where an initial RMSE of 77.5 meters was reduced to 41.5 meters with just 14 measurements. Finally, georeferencing with accurate intrinsic and extrinsic parameters, as was the case in the UAVision experimental procedure, resulted in very accurate estimated positions. An average RMSE of 16.33 meters was

achieved with a total of 18 landmarks that distanced more than 1000 meters from the UAV.

VI-A. Future Work

Regarding the acquisition of multiple images of the same target, the next step would be to design flight patterns that minimize the error of the IRT algorithm but also maximize the information extracted from Bearings-Range measurement model. In addition, considering the last experimental procedure, an algorithm could be developed that searches recognizable landmarks in the imagery acquired by the aerial vehicle. These landmarks could then be georeferenced using image registration techniques and used to refine the camera's EPs.

REFERENCES

[1] EU-DEM v1.1. URL https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=metadata. visited on 26-07-2020.

[2] D. Blake Barber, Joshua D. Redding, Timothy W. McLain, Randal W. Beard, and Clark N. Taylor. Vision-based target geo-location using a fixed-wing miniature air vehicle. *Journal of Intelligent and Robotic Systems: Theory and Applications*, 2006.

[3] Gianpaolo Conte, Maria Hempel, Piotr Rudol, David Lundström, Simone Duranti, Mariusz Wzorek, and Patrick Doherty. High accuracy ground target geo-location using autonomous micro aerial vehicle platforms. *AIAA Guidance, Navigation and Control Conference and Exhibit*, 2008.

[4] Davide Antonio Cucci, Martin Rehak, and Jan Skaloud. Bundle adjustment with raw inertial observations in UAV applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2017.

[5] Gianfranco Forlani, Fabrizio Diotri, Umberto Morra di Cella, and Riccardo Roncella. Indirect UAV Strip Georeferencing by On-Board GNSS Data under Poor Satellite Coverage. *Remote Sensing*, 2019.

[6] Mehran Ghandehari, Barbara P Buttenfield, and Carson J Q Farmer. Comparing the accuracy of estimated terrain elevations across spatial resolution. *International Journal of Remote Sensing*, 2019.

[7] Fangning He, Tian Zhou, Weifeng Xiong, Seyyed Meghdad Hasheminnasab, and Ayman Habib. Automated aerial triangulation for UAV-based mapping. *Remote Sensing*, 2018.

[8] Christian Heipke, Krasten Jacobsen, H Wegmann, and Birgir Nilsen. Integrated sensor orientation-an OEEPE Test. *International Archives of the Photogrammetry and Remote Sensing*, 2000.

[9] Elder M. Hemerly. Automatic georeferencing of images acquired by UAV's. *International Journal of Automation and Computing*, 2014.

[10] Dong Wook Kim, Hee Sup Yun, Sang Jin Jeong, Young Seok Kwon, Suk Gu Kim, Won Suk Lee, and Hak Jin Kim. Modeling and testing of growth status for Chinese cabbage and white radish with UAV-based RGB imagery. *Remote Sensing*, 2018.

[11] Frederik S. Leira, Kenan Trnka, Thor I. Fossen, and Tor Arne Johansen. A ligth-weight thermal camera payload with georeferencing capabilities for small fixed-wing UAVs. *2015 International Conference on Unmanned Aircraft Systems, ICUAS 2015*, 2015.

[12] A. M.G. Lopes, A. C.M. Sousa, and D. X. Viegas. Numerical simulation of turbulent flow and fire propagation in complex topography. *Numerical Heat Transfer; Part A: Applications*, 1995.

[13] Sameera S. Ponda, Richard M. Kolacinski, and Emilio Frazzoli. Trajectory optimization for target localization using small unmanned aerial vehicles. *AIAA Guidance, Navigation, and Control Conference and Exhibit*, 2009.

[14] Yongwei Sheng. Comparative evaluation of iterative and non-iterative methods to ground coordinate determination from single aerial images. *Computers and Geosciences*, 2004.

[15] Chen Xiong, Qiangsheng Li, and Xinzheng Lu. Automated regional seismic damage assessment of buildings using an unmanned aerial vehicle and a convolutional neural network. *Automation in Construction*, 2020.

[16] Cheng Xu, Daqing Huang, and Jianye Liu. Target location of unmanned aerial vehicles based on the electro-optical stabilization and tracking platform. *Measurement*, 2019.

[17] Zhengyou Zhang. Flexible camera calibration by viewing a plane from unknown orientations. *Proceedings of the IEEE International Conference on Computer Vision*, 1999.