

VISION BASED NAVIGATION FOR PROXIMITY OPERATIONS AROUND ASTEROID 99942 APOPHIS

Michel Delpech⁽¹⁾, Vincent Bissonnette⁽²⁾, and Laurent Rastel⁽³⁾

⁽¹⁾⁽³⁾ CNES, 18 Avenue Edouard Belin, 31400 Toulouse, michel.delpech@cnes.fr

⁽²⁾ ISAE, 18 Avenue Edouard Belin, 31400 Toulouse, vincent.bissonnette@isae.fr

Abstract: *The work presented in the paper concerns the analysis of monocular vision based navigation techniques enabling autonomous operations around an asteroid at close range to characterize its gravity field. The paper focuses on the state estimation layer and performs a comparison of two different SLAM methods (EKF-SLAM and Sparse Bundle Adjustment) from accuracy, computational cost and robustness points of view. This work continues with a parametric sensitivity analysis for a generic orbiting scenario. In the last stage, the inclusion of the vision layer in the navigation system is analyzed and evaluated. Current results are synthesized and discussed.*

Keywords: *Vision based navigation, SLAM, Bundle Adjustment, EKF, asteroid*

1. Introduction

In recent years, there has been a growing interest in the space community for the Near Earth Objects and particularly the Potentially Hazardous Asteroids (PHA) which possible impact with Earth could be catastrophic given their size. CNES initiated in 2012 a mission concept study devoted to the scientific analysis of the asteroid 99942 Apophis in the perspective of its close encounter with Earth in April 2029. Among the various objectives that included the in-situ analysis on the surface with a dedicated lander, the determination of the asteroid gravity field added strong requirements on the vehicle maneuvering autonomy, in particular the capability to navigate for a long duration over the asteroid at very close range.

So far, all missions to asteroids or comets have relied on some extensive involvement of ground analysts to navigate the probe. In missions like JAXA's Hayabusa [1] and ESA's Rosetta[2], the man-in-the-loop process was concerning the mapping of the celestial object surface and included also the probe localization. Lately, NASA missions like Near Shoemaker [3] and Dawn [4] have introduced more autonomy on-board by implementing automatic landmark recognition techniques but the lengthy and tedious task of mapping the asteroid remained on the ground. For future and more challenging missions, the desire to reduce the operator load and increase the navigation robustness in case of tracking loss leads to the use of techniques that can perform at the same time localization and mapping of landmarks that are detected in the field of view.

This trend can be supported by recent developments in the Simultaneous Localization And Mapping (SLAM) domain that have triggered numerous implementations in the robotics field for both wheeled mobile vehicles and UAVs. So far, the most popular SLAM approach uses a standard Extended Kalman Filter to estimate recursively the vehicle state and the location of visual features. More recently, the Structure From Motion (SfM) approach that relies initially on global optimization has been adapted into efficient sequential methods that perform sparse Bundle Adjustment (BA). Some studies showed also that the sequential bundle adjustment

(BA-SLAM) approach could outperform EKF-SLAM in accuracy at a reasonable computational cost.

The purpose of this paper is to compare the relative merits of these two SLAM techniques for the specific scenario of asteroid orbiting and investigate possible hybridation solutions. First, the analysis is focused on state estimation assuming particular vision performances and relies on some evaluation framework specially designed to emulate various navigation scenarios and perform parametric study. In a second stage, a SURF feature detection and tracking functionality is integrated in the navigation process to confirm the validity of the analysis results in presence of simulated asteroid images and illustrate the system behavior performance.

The paper is organized as follows: in Section 2, the mission concept, the technical background and the scope of the problem are explicated. After a presentation of the navigation assumptions, Section 3 describes the implementation of the two candidate SLAM methods and details their comparative analysis on a test case that is representative of orbiting scenarios. Next, Section 4 is devoted to the assessment of the complete navigation system combining the image processing and state estimation layers : the experimental testbed and scenario are detailed, the results are described and discussed. Finally, the main lessons are summarized in a conclusion where current limitations and perspectives for further work are presented.

2. Background

2.1 Reference mission concept

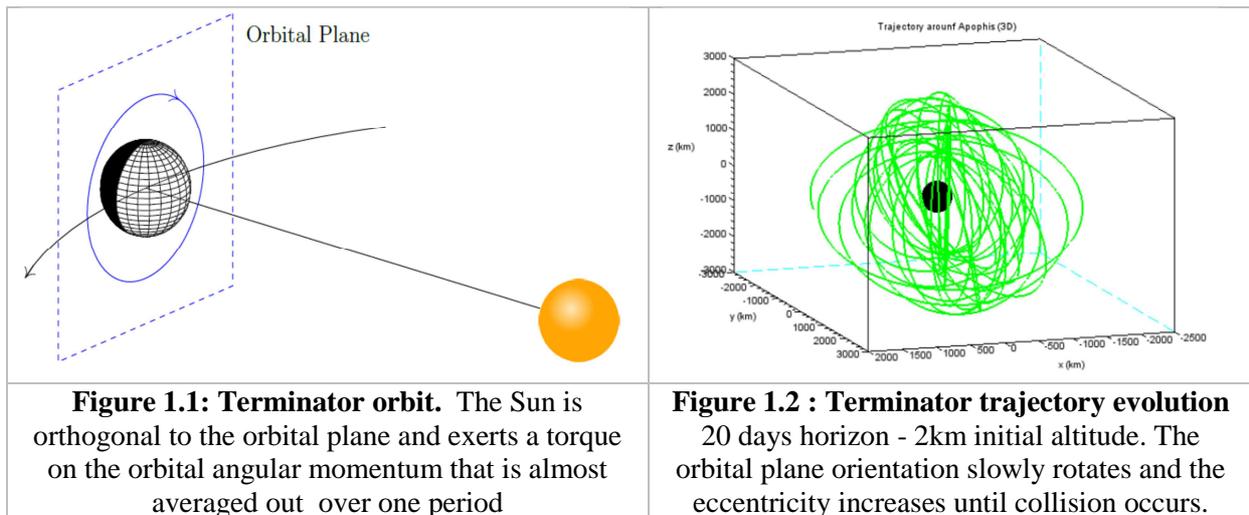
Apophis 99942 discovered in 2004 belongs to the Potential Hazardous Asteroid population that is steadily growing and must be monitored with close attention through ground observation means. The Apophis Earth flyby in 2029 will thus offer a great opportunity to acquire a better knowledge of this type of objects and investigate the feasibility of mitigation techniques. For that purpose, CNES initiated in 2012 a mission concept study involving a rendezvous with the asteroid to analyze its interior structure using seismic or radar sounding techniques [5]. The different scientific scenarios required operations at medium to close range (down to a few hundreds of meters) and the most challenging one concerned the characterization of the gravity field that was not compatible with flyby or hovering trajectories. The mission concept study was actually terminated in 2014 but internal R&D work was pursued at CNES in the guidance and navigation domain to assess the technical feasibility of the autonomous orbiting operations for gravimetry.

Table 1 : Apophis characteristics

Characteristics	Value	Comment
Diameter	375 +/- 15 m	2014 observations
Mass	$4.4 < m < 6.2 \cdot 10^{10}$ kg	Based on geometrical model
Spin	30.4 h	Tumbling motion
Albedo geometrical	0,30 +/- 0.06	Herschell data
Aphelion / perihelion	1.0985 AU / 0.7460 AU	

Achieving the gravimetry of Apophis constitutes a tough challenge considering its low gravity field. In previous missions such as Near Shoemaker or Dawn, the gravimetry has been performed around objects with diameters ranging from tens to hundreds of kilometers (35 km for Eros, 530 km and 950 km for Vesta and Ceres). The body attraction was therefore large enough compared to the others perturbations to place the spacecraft on some orbital

trajectories and ensure long duration periods with no actuation. In the case of small bodies like Apophis, the impact of solar radiation pressure becomes prominent and limits drastically the capability to find quasi-stable orbits. It remains possible anyway to limit this impact by inserting the spacecraft on a low terminator orbit (Figure 1.a) that benefits from some cancellation effect. Preliminary studies have shown that actuation free trajectories could be achieved for weeks with a low collision risk assuming initial altitudes around 2 km (Figure 1.b). However, the lower the altitude, the higher the risk of collision for a given duration. Considering the gravimetry accuracy requirements, the orbit altitude should be in the 500 m range which implies some spacecraft autonomy at the navigation and guidance levels to maximize the free trajectory duration in a safe manner. The paper will focus on the navigation aspects.



2.2 Navigation

The major difficulty when navigating around small bodies comes from the need to achieve localization within a poorly known environment that requires some extensive mapping. In the most ambitious missions flown so far, the two processes of mapping and localization have been separated. First, using numerous images of the body, the ground analysts perform the tedious task of selecting landmarks and building a database that is uploaded to the spacecraft. In a second phase, the spacecraft can use these landmarks as navigation tie-points by running automatic landmark recognition and tracking algorithms. NEAR Shoemaker that orbited and landed on Eros in 2001 relied successfully on a database of craters (images from multiple points of view) that constituted the prominent population of features on this asteroid. Later, more advanced landmark descriptors were developed for the Dawn mission that performed low altitude orbiting around Vesta in 2011 and Ceres in 2015: the local region surrounding any distinct feature is represented by a landmark map (L-map) that captures both elevation and albedo information. Using the L-maps, reference landmark images can then be rendered on-board and compared with real images. This approach allows improving the robustness of the tracking process but the building of any L-map requires a large collection of images of the same landmark with different lighting conditions.

Getting rid of the operator while achieving at least the same level of robustness constitutes a major challenge that has been addressed in multiple studies [6,7]. EKF-SLAM has been the preferred method so far for computing considerations and has been evaluated in conjunction with various feature detection and tracking algorithms. Since automatically detected features are more sensitive than L-maps to viewpoint change and incident light variations, the typical

solution to achieve robustness is to take benefit from statistical effects by the processing of a large number of features. These circumstances may prove more favorable to alternative techniques such as BA-SLAM and it is therefore relevant to analyze the respective pros and cons of these two methods in some parametric way for the particular context of asteroid orbiting. The main characteristics of these techniques are briefly discussed in the sequel.

SLAM filtering: The EKF-SLAM method that is known to be computationally efficient for small number of features presents a major advantage: the direct availability of the camera and landmark position uncertainty from the state covariance matrix. First, this information helps to detect anomalies such as filter divergence and allows triggering corrective action if needed. In addition, when the filter is coupled to some vision system, the uncertainty measure enables to reduce the search region during the features tracking process which robustness and computing efficiency can then be enhanced. The general practice to achieve accuracy and robustness consists in using all possible a priori information on the camera dynamics as well as inertial sensor measurements to predict the camera state for each image acquisition [8]. Some implementations however manage to get satisfactory estimation results with a motion modelled as a Gaussian random walk [9].

For monocular vision, EKF-SLAM has to manage new landmarks whose distance uncertainty cannot be properly represented as a gaussian distribution. The widely accepted solution to tackle this problem is the Anchored Homogeneous Point parametrization [10] that introduces the landmark inverse depth in the state vector. This converts however in a larger vector size that carries 6 variables for each landmark instead of 3 for a Euclidean point. The current technique to limit the corresponding memory usage and computing complexity consists in switching to the Euclidean representation after a given reduction of the distance uncertainty. Another EKF-SLAM weakness comes from its rather high sensitivity to measurement errors due to spurious matches of landmarks in consecutive images. Some improvement of the filter robustness can be achieved by different methods such as the *active search* that limits feature extraction in regions where they are more likely to be found [11], the *Joint Compatibility Branch and Bound Test* [12], and the *1-Point RANSAC* algorithm [13].

BA-SLAM: The original BA represents a very efficient method to estimate the location of landmarks and the camera by processing numerous observations in a batch mode. While this implementation operates on the complete graph that links all observation points to all the visible landmarks, the sparse BA processes data from only a subset of views at each estimation step. To achieve computing efficiency, a solution is to perform estimation on a sliding window that captures information from a given spatial or temporal horizon without any consideration of the measurements value. Some enhancement of this method consists in selecting the number of views that conserve the least possible redundant information and various studies have attacked this problem [14]. The main advantage of the BA method that processes measurements from various viewpoints is its higher accuracy and better robustness to spurious matches that are being averaged out. In addition, BA that relies on non linear optimization through Gauss Newton methods is less sensitive to errors in the predicted camera motion and some implementations simply assume no motion between two consecutive observations [15]. Conversely, a possible weakness of BA w.r.t. EKF-SLAM is the lack of explicit representation of the estimated state uncertainty even though gaussian random variables are actually modelled in the process. Getting access to the state covariance matrix adds therefore a significant computational overhead that can be an impediment to the application of landmark active search methods. Another limitation comes forward in monocular navigation: the camera and landmark depth information is only known with a scale ambiguity factor unless some known object can be identified in the scene and used as depth

reference. Finally, BA is also subject to drift since the reference object cannot be kept forever in the number of views used in the estimation process. New references within the current visible landmark population need to be initialized using depth estimated information and this will induce some error increase over time. It has been shown in a previous study [16] that the computational complexities of the EKF-SLAM and BA-SLAM are respectively $O(N^2)$ and $O(NM^2 + M^3)$, with N being the number of processed features and M the camera views number. In addition, this study stated that sequential BA could outperform EKF-SLAM in a large variety of situations since it can make a better use of the computation capability in terms of accuracy.

Vision layer: SLAM approach credibility is tightly correlated to the vision layer ability to detect and track features robustly and with an acceptable computational cost. In the computer vision domain, new algorithms for automatic feature detection and tracking have emerged and show some enhanced robustness in presence of image transformations and variations of lighting conditions. These algorithms belong to the group of Local Invariant Feature detectors and the most popular ones in the literature are SIFT, SURF and ORB [17,18,19]. These algorithms allow to detect some distinctive image features (edges, corners, blobs, ..) and build some synthetic representation of the local neighbourhood that is stored in a so-called descriptor. To achieve some invariance to viewpoint changes and facilitate the feature matching process, the feature model assumes that the local region is planar and this constitutes of course the technique main limitation when areas with significant relief are observed. The matching process relies on the comparison of descriptors that have been designed to be the most distinctive and regarded as potential IDs: the process consists in searching for each feature in image A what is the candidate from image B providing the minimum descriptor difference (score). To reduce the probability of returning false matches, Lowe [20] proposed some constraint to be satisfied: a candidate is selected if the ratio between its score and the second best remains below a given threshold. In addition, the resulting matches can be filtered through the application of a RANSAC algorithm that allows rejecting outliers assuming some image homographic transformation.

3. Description and evaluation of state estimation methods

3.1. Problem description:

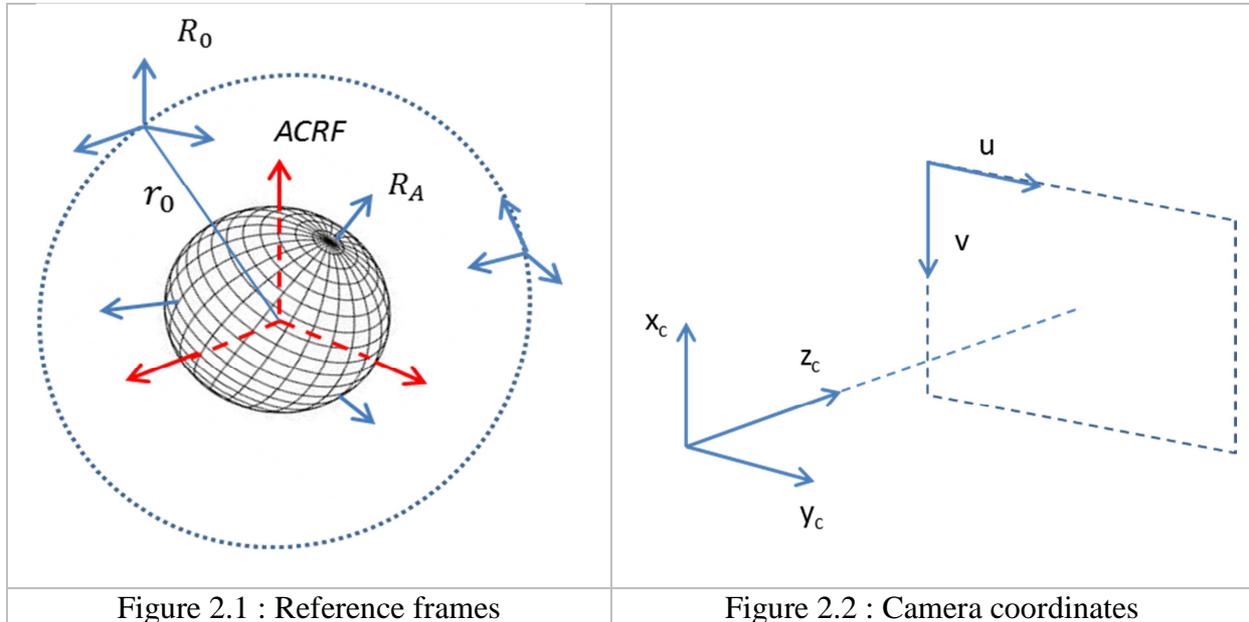
The representation of the different motions require the definition of the following reference frames (see Figure 2.1)

- Asteroid Centered inertial Reference Frame $ACRF$: J2000 frame translated to the asteroid assumed center of mass
- Camera frame R_C : z axis aligned with the camera boresight, x axis along the sensor vertical and y axis along the horizontal (see Figure 2.2)
- Initial location reference frame R_0 : J2000 frame translated to the camera origin
- Asteroid body frame R_A : origin at the assumed center of mass (axes definition indifferent)

The reference scenario considered for analysis can be summarized by the following set of assumptions:

- The probe is flying around the asteroid on a circular orbit
- The asteroid motion is perfectly known

- The initial position of the spacecraft (SC) is perfectly known in ACRF at the first frame
- the asteroid is assumed spherical with an isotropic distribution of landmarks
- the landmarks visibility is taken into account considering the relative landmark-camera geometry and the camera field of view
- the vision layer is simulated by introducing a certain rate of landmark disappearance that is proportional to the amplitude of the viewpoint change



3.2. Description of the BA-SLAM implementation

In this section we are considering the version of the BA-SLAM that processes at each iteration step a constant number of views. The estimation concerns two types of information: the camera poses associated to the different views and the positions of landmarks that have been observed in at least two consecutive views. The state vector x is represented as follows:

$$x = [l_1, \dots, l_N, T_1, \dots, T_m]^T \quad (1)$$

with vectors l_i representing cartesian landmark positions expressed in the initial reference frame and camera poses T_j composed of a vector of 3D cartesian coordinates in the initial reference frame and a quaternion indicating orientation with respect to the initial reference frame. As regards the observation vector y , each one of the p landmarks provides 2 bearing angles only since the vision is monocular. The observation vector dimension is therefore $2p$ with $p < mN$ as the landmarks are observed multiple times.

The optimization task consists in finding the state x that represents the best fit to the observations. The cost function takes the form of a scalar product of an error vector function :

$$c(x) = (y - f(x))^T W_y (y - f(x)) \quad (2)$$

where $f(x)$ is the reprojection vector of landmarks in the camera frames corresponding to a given state x and W_f is the weight matrix taking into account the measurement uncertainty.

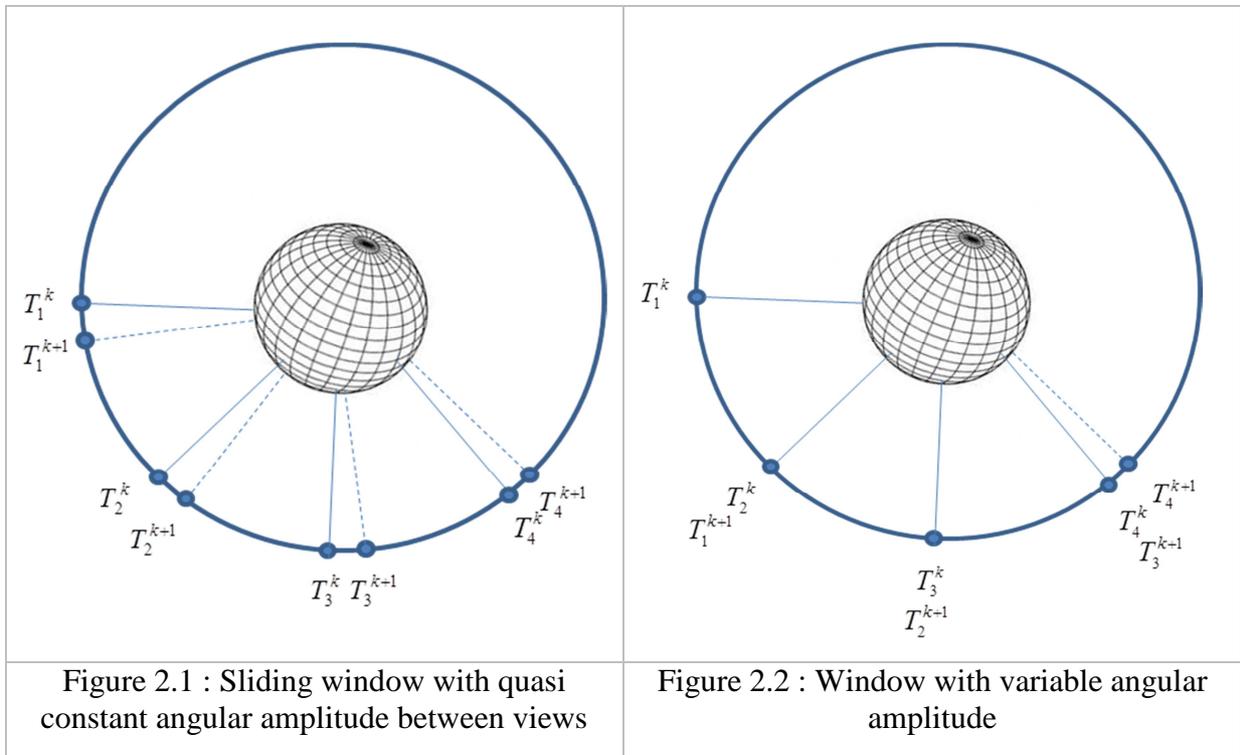
Since $c(x)$ is a quadratic function that approximates the null value when reaching its minimum, iterative Gauss-Newton optimisation is actually applicable. This method consists in computing at each iteration the state increment that solves the following equation:

$$J_f(x)^T W_f J_f(x) dx = -J_f(x)^T W_f f(x) \quad (3)$$

where $J_f(x)$ is the jacobian matrix of the reprojection vector $f(x)$. To improve convergence, our implementation relies on some efficient adaptation of this method named Levenberg-Marquard (LM) [20]. The equation to be solved has an augmented matrix including a damping term λ which is scaled to ensure cost reduction while keeping the matrix semi positive :

$$(J_f(x)^T W_f J_f(x) + \lambda I) dx = -J_f(x)^T W_f f(x) \quad (4)$$

Keyframe management: The principle of the method is to restart the optimization process each time a new image is acquired. To limit the algorithm complexity, it is therefore required to use a limited number of frames that compose a so-called sliding window. The process consists in managing a graph that stores the current camera pose and a fixed number of poses acquired in the past. All the other frames and observations are ignored in the optimization process but stored for later use. The selection of the right keyframes constitutes here a key issue. We tend to favor large intervals between consecutive views since it yields the benefit of a large observation base. Conversely, it reduces the number of landmarks being observed in multiple views which affects the estimation performance. It is therefore understood that a tradeoff value depending on the landmark distribution can be found and this is one of the purposes of the analysis performed in section 4.



Two frame management strategies have been considered. Here, we focus on the steady mode and we do not detail the window initial construction that is straightforward. For the sake of clarity we assume that the number of frames being processed at every step is equal to 4.

The first strategy illustrated on Figure 2.1 consists in keeping a constant angular offset between consecutive views. Whenever a new *leading* frame T_4^{k+1} is acquired after a certain camera displacement, another set of *trailing* frames $\{T_1^{k+1}, T_2^{k+1}, T_3^{k+1}\}$ is selected in the stored map to produce increments of identical amplitude. Since the observation base is maintained quasi constant and close to the optimum value, the estimation performance can be maximized.

In the second strategy shown on Figure 2.2, the window structure is only modified by the shift of the leading frame since the trailing frames are generally kept the same. Periodically, the whole window structure needs to be updated when the distance between the two last frames T_3^k and T_4^k cannot ensure a proper landmark connectivity. When this occurs, the first window frame is lost and the new sequence is obtained by shifting the frame number as follows: $T_1^{k+1} = T_2^k$, $T_2^{k+1} = T_3^k$, $T_3^{k+1} = T_4^k$. The window size gets therefore reduced before growing again and the estimation efficiency is likely to be impacted accordingly. However, this strategy has the important merit of sparing memory with respect to the previous one since the useful data history is always kept in the window frames. Benefiting also from the advantage of simplicity, this method has been chosen for our implementation.

Camera motion : The BA-SLAM algorithm can be used without any motion model and the method in that case consists in initializing the optimization process by setting the new pose equal to the previous one. This is the typical method in photogrammetry applications where the knowledge of the camera velocity can be either ignored or reconstructed a posteriori via data smoothing techniques. In our context, the velocity information is needed to perform on-board guidance and it must be reconstructed by some independent estimator. The most efficient technique consists in using a dynamic model of the vehicle that integrates the asteroid attraction and the solar radiation pressure as it is done in a typical navigation system. This model can be therefore used in a propagator that will provide the predicted position of the next frame. To keep the estimated velocity accurate, the predicted position needs to be updated using the output of the BA-SLAM algorithm according to a scheme further detailed at the end of the section.

BA-SLAM steps : In order to achieve proper convergence in the most difficult cases and particularly when there is a low quality motion model or no model at all, it is common to divide the bundle adjustment process in three optimization steps which progressively update the graph. The process being carried out at each new image frame acquisition is the following:

Preparation phase:

- Extract landmark measurements;
 - Set new camera pose according to the motion model and append to state vector;
 - Manage camera poses and views according to the selected strategy – this step involves the verification the graph connectivity is adequate (sufficient landmark overlap between views)
1. **Perform Motion-Only BA:** An adjustment is performed over the current camera pose only, using current measurements and current best estimation of landmark states. This step effectively serves the purpose of a motion model;
 2. **Perform Structure-Only BA:** An adjustment of the landmark positions only is performed using information from the free views;
 3. **Perform Full BA:** Using state estimations of the two previous steps, a full joint optimization over the sparse graph is performed. Assuming the motion and structure updates have succeeded, the state should already be close to the minimum, and convergence shall be reached in relatively few steps.

In presence of a good quality motion model, steps 1 and 2 can be skipped to reduce the computational complexity while preserving the estimator performance.

Depth gauge management: To tackle the scale ambiguity problem, the depth of one of the initial frame landmarks is assumed to be known and is used afterwards as a gauge. For that purpose, the retroprojection error vector from equation (2) is augmented with a quantity representing the difference between the predicted distance of this particular landmark noted as l_1 and the gauge value as observed in frame T_1 :

$$y - f(x) = \begin{pmatrix} y_l - f_l(x) \\ d - \hat{d}(l_1, T_1, x) \end{pmatrix} \quad (5)$$

This gauge from frame T_1 can be used in the optimization process as long as the connectivity with the last acquired view can be maintained. With a fixed number of views, the connectivity is lost when the camera displacement reach a limit value that depends directly on the asteroid local aspect and the feature detection and tracking performance. When this limit is reached, it is necessary to select a new landmark which depth will be considered as reference. Since this information has been obtained through the estimation process and carries some error, a depth bias is introduced in the subsequent optimization steps and will propagate until the next gauge change. In absence of range measurement or external update, the depth error is due to drift with a behaviour similar to some random walk process.

Motion propagator update: After each BA-SLAM iteration step, a new estimate of the current camera position is available and this information is used in our implementation to update the state of the camera motion estimator running in parallel. Assuming a perfect camera attitude knowledge, a simple 6 states EKF filter can be used to estimate the position and velocity expressed in the ACRF asteroid inertial frame. The dynamic model is the following one:

$$\ddot{r} = \frac{\mu}{r^3} r + v_{srp} + \zeta \quad (6)$$

where r and \ddot{r} are the position and acceleration vectors, μ the assumed asteroid gravitational constant, v_{srp} the acceleration due to solar radiation pressure and ζ sums up all the non modelled forces. Considering the high cost of computing the camera position uncertainty, we implement a filter with a constant Kalman gain [19].

3.3. Description of the EKF-SLAM implementation

With respect to BA-SLAM, the EKF carries a lighter state vector since it ignores the previous camera poses along with the landmarks not visible any more in the current image:

$$x = [X_C, l_1, \dots, l_N]^T \quad (7)$$

In general, the current camera state X_C is represented by a 13-element vector composed of the 6 position and velocity coordinates expressed in the ACRF frame, the attitude quaternion with respect to the same frame and the angular rate expressed in the camera frame.

When using monocular vision, the absence of distance measurement implies to use a landmark state representation that does not affect negatively the filter convergence. Linearity is therefore paramount and the Anchored Modified Polar Point representation approaches this objective by defining a landmark with the following 6-element vector:

$$l_i = [u, v, \rho, T_0] \quad (8)$$

T_0 is referred as the anchor and represents the camera position in the ACRF frame when the landmark is added in the state vector. The u, v parameters are the landmark pixel coordinates in the camera frame when observed from the anchor position and the parameter ρ is the inverse of the distance between the landmark and the anchor position. Using an inverse distance parameter allows to represent within bounds any potentially unlimited uncertainty and preserve also the Gaussian character of the measurement distribution.

The EKF-SLAM sequential process includes the two typical steps encountered in any Extended Kalman Filter: (1) State Prediction and Covariance Propagation using some motion model, (2) State and Covariance Update when new measurements get available. In addition, the EKF-SLAM process includes a third and specific operation corresponding to State Management: this operation includes the addition of new landmarks in the state vector as well as the removal of landmarks no more visible – it involves also the associated modification of the covariance.

State Management: Throughout the orbiting of the asteroid, new landmarks keep appearing in the camera field of view and have to be considered by the filter in order to pursue the mapping and preserve the camera position observability. The first step of the landmark initialization consists in producing an augmented state vector by appending the landmark 10-element vector to the current state:

$$\hat{x}_{aug} = [\hat{x}, u, v, \rho, T] \quad (9)$$

where ρ represents an initial guess of the landmark inverse distance and T is the estimated camera pose with 7 elements.

Since the landmark distance can typically take any value from 0 to the distance of the asteroid surface, this parameter can be considered part of the estimator configuration along with its associated uncertainty R_ρ that is chosen large enough to guarantee its later convergence to the true value. The landmark state covariance P_{ll} is then initialized according to the following formulas :

$$P_{ll} = \frac{\partial l}{\partial X_c} P_{XX} \frac{\partial l}{\partial X_c}^T + R_l \quad (10)$$

$$R_l = \begin{bmatrix} R_y & 0 & 0_{3 \times 3} \\ 0 & R_\rho & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \end{bmatrix} \quad (11)$$

where P_{XX} represents the covariance of the current camera state, $\frac{\partial l}{\partial X_c}$ is the partial derivate of the landmark state l with respect to the camera state X_c and R_y is the direction measurement uncertainty.

Gauge Fixing and the Scale Problem: With monocular vision, the EKF-based SLAM estimator faces the same scale ambiguity problem as BA-SLAM does. In our implementation, we apply therefore a similar approach to fix the gauge at the start of the estimation. This consists in assuming that the range of one of the landmarks visible in the first image can be measured with sufficient accuracy. To that purpose, the initial guessed parameter ρ_0 of Equation (9) is set with the measured inverse range $\rho^* = 1/d$ and the covariance component

R_ρ in R_l is set with its associated small uncertainty value. In subsequent time steps, the propagation of this precise information to the new camera pose and landmark states allows to keep the scale constrained.

4. Comparative evaluation of the SLAM methods

These two methods have been implemented in the Matlab environment for the quick prototyping capability and to benefit particularly from the Optimization toolbox. Considering that both methods are prone to drift, the evaluation will focus on the evolution of the estimation error over time. To perform a relevant method comparison, we are considering a reference estimation scenario that is detailed hereafter:

- the motion around the asteroid is circular and the difference of anomaly between all consecutive image frames is constant : Δa
- the number of landmarks detected in every image frame is constant : N^1
- the landmark ‘repeatability ratio’ for the Δa angular increment is constant : k^2
- the motion prediction error between two consecutive image frames is constant and proportional to the displacement : λ
- the bearing measurement error is constant for all landmarks (3σ) : ϵ_m
- the initial position is assumed perfectly known
- the camera attitude is assumed perfectly known as well as the camera parameters
- the number of views considered in the BA-SLAM process : M

- 1: The landmark population is uniform on the asteroid and the density is adapted to provide in average a specific number of landmarks per frame N
- 2: Considering a group of N landmarks detected in frame k and visible in frame $k+1$, the ‘repeatability ratio’ is the number of landmarks successfully tracked in frame $k+1$ divided by N

4.1 BA-SLAM behaviour

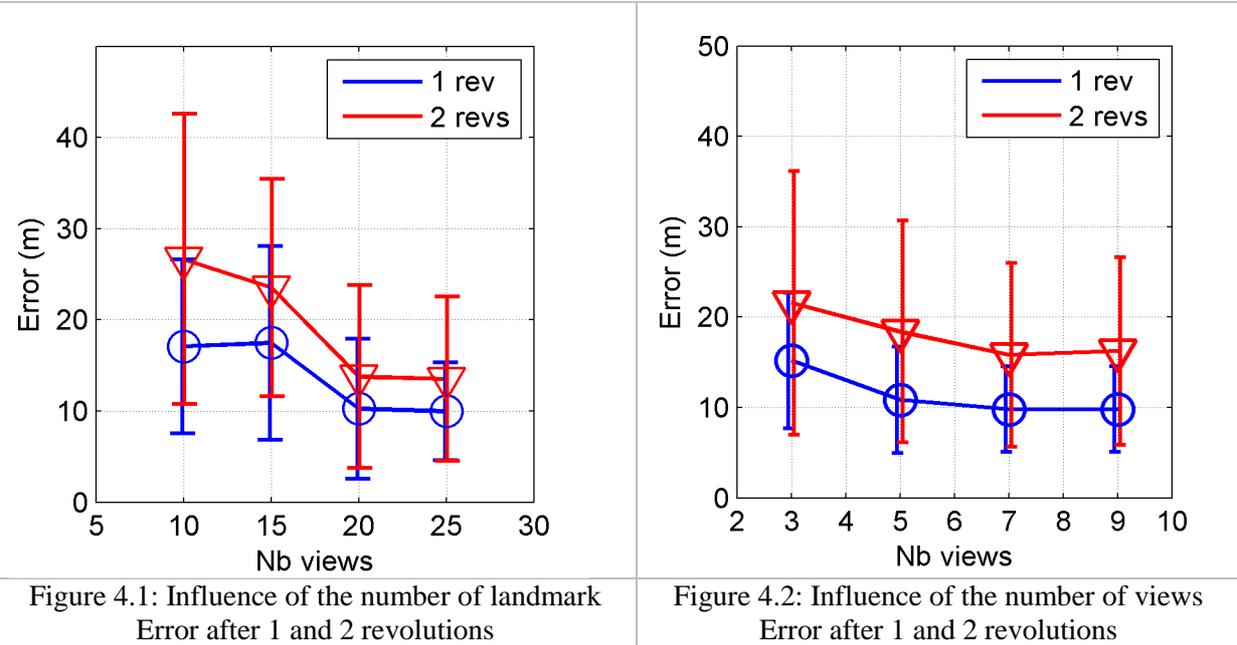
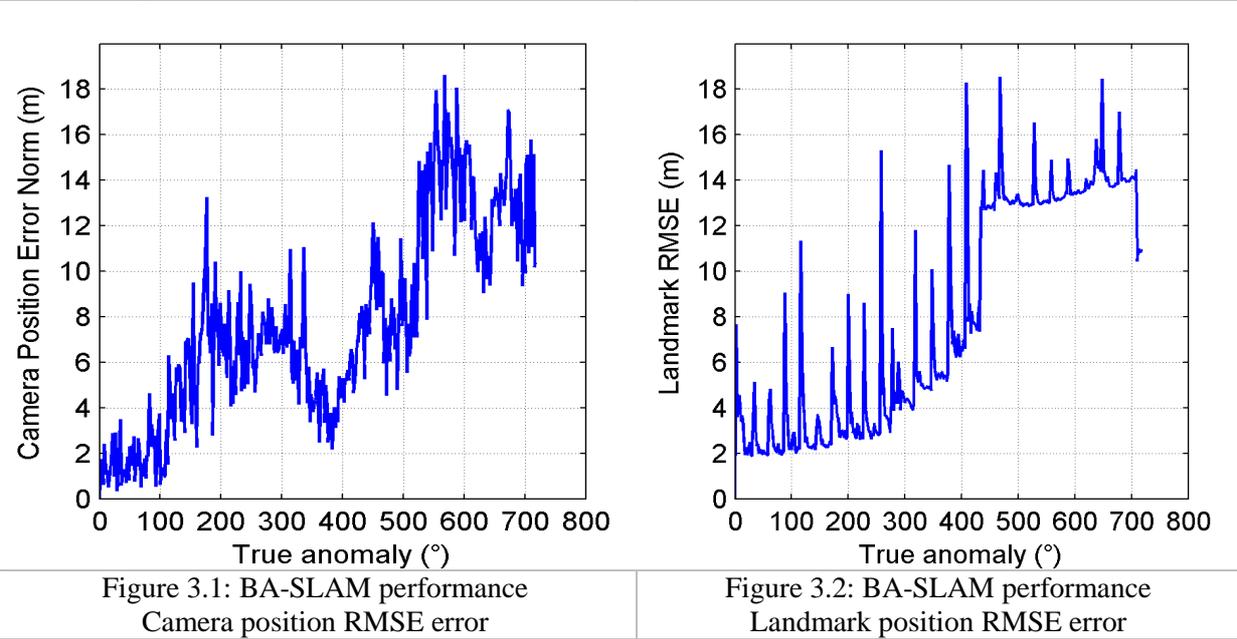
The analysis starts with a characterization of the BA-SLAM method focused on its sensitivity to the most critical parameters: the number of views (M) and the number of landmarks detected per view (N). The performance criteria considered are the camera position errors after 1 and 2 revolutions (1 revolution lasts about one day). The test conditions are summarized on Table 2:

Table 2 : BA-SLAM characterization

Parameter	Value / test 1	Value / test 2
Number of landmarks/view : N	[10, 15, 20, 25]	15
Displacement /view : Δa	2°	2°
Landmark repeatability : k	0.90	0.90
Motion prediction : λ	0.5	0.5
Bearing measurement error : ϵ_m	2 pixels	2 pixels
Number of views : M	5	[3, 5, 7, 9]
Max between 2 consecutive views	30°	20°
Minimum connectivity between views	3 landmarks	3 landmarks

Due to the dispersion of results, 100 Monte Carlo simulations are run per test case and performance focusing on the camera position estimation is summarized by the mean error

value and error standard deviation after 1 and 2 revolutions. The BA-SLAM behaviour is illustrated on Figures 3.1 and 3.2 in a simulation run with 5 views and 15 landmarks: they show the evolution of the camera position error norm and the landmark position RMS error (only the landmarks processed at the current step are considered). The effect of the initialization of new landmarks can be clearly observed by the sudden increase of the landmark position error.

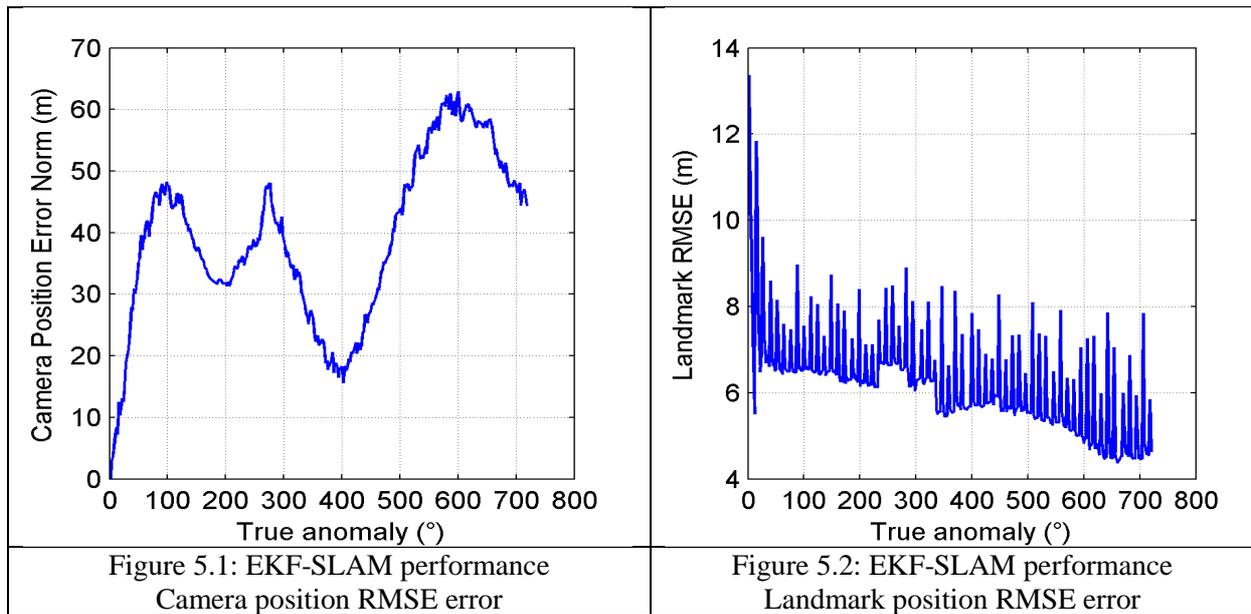


The benefit of tracking a larger number of landmarks is illustrated on Figure 4.1 with a 60% reduction of the position error after 1 or 2 revolutions when going from 10 to 20 landmarks. The improvement is barely noticeable going up to 25 landmarks where only a reduction of the dispersion can be observed. Similarly, increasing the number of views from 3 to 7 provides a 40% reduction of the position error as illustrated on Figure 4.2 but increasing it any further does not bring any improvement. A thorough analysis is currently underway to better

understand why some performance stall is observed for a certain number of landmarks and views.

4.2 EKF-SLAM behavior

The EKF-SLAM behaviour is illustrated on Figures 5.1 and 5.2 that show the evolution of the camera position error and the landmark position RMS error. The configuration is done according to Table 2 general parameters (except specific BA-SLAM ones) and with 15 landmarks per view. At the beginning, the uncertainty is generally increasing faster than BA-SLAM but is reduced afterwards thanks to the filter convergence.



4.3 Performance comparison

The comparison is performed between EKF-SLAM and the BA-SLAM implementation with 5 views. We consider 3 cases corresponding to different numbers of landmarks per view (10, 15 and 20) and 100 Monte Carlo simulations are run per case. Performance metrics include the mean cycle execution time, the RMS error for all the landmarks observed, the norm of the camera position error after 1 revolution. For reference, simulations were performed in a 32 bit Matlab environment running on a 2.40 GHz processor.

Table 3: Comparison synthesis

Case	Metric	EKF-SLAM	BA-SLAM (5 views)
10 landmarks	Cycle Execution time (s)	0.020	0.077
	Landmark RMSE (m)	3.72 +/- 1.9	8.33 +/- 2.2
	Camera position error norm (m)	25.0 +/- 25.8	17.08 +/- 9.5
15 landmarks	Cycle Execution time (s)	0.049	0.118
	Landmark RMSE (m)	2.72 +/- 2.1	7.21 +/- 2.6
	Camera position error norm (m)	24.07 +/- 20.0	17.42 +/- 10.3
20 landmarks	Cycle Execution time (s)	0.119	0.158
	Landmark RMSE (m)	1.45 +/- 1.8	5.48 +/- 1.47
	Camera position error norm (m)	8.98 +/- 5.4	10.25 +/- 7.6

As predicted, the EKF-SLAM implementation is less computationally-intensive than BA-SLAM but this advantage gets smaller when the number of landmarks increases. This comes as expected at the price of a lower accuracy and particularly when the number of landmarks per view is less than 20. At this stage, the execution time metric must be considered carefully since the cost of the image processing was not accounted for. This topic is addressed in next Section.

5. Experiment

In the previous section, both SLAM algorithms have been characterized using simulated landmark measurements. This type of data was produced by setting specific values to parameters like the feature distribution on the surface, the level of noise in pixel coordinates, and the percentage of tracked features between views. The purpose of this experiment is to improve the analysis relevance through the addition of the feature detection and tracking algorithms into the system.

Image Generation: To simulate Apophis images, we used a powerful open source product: the 3D computer graphics Blender software [20]. This product provides a convenient API in Python language to design the scene, control the position of the different objects, and export screenshots of the field of view. A script has been developed to recreate the trajectory of our reference scenario, export the field of view as an image at every time step, and export a text file containing the true camera pose information and relevant simulation parameters for further analysis. Figure 4.1 shows a typical output image, here obtained at an altitude of 600 m and a field of view of $30^\circ \times 30^\circ$. The shape model of asteroid Apophis was simplified down to a sphere of 190 m radius, on which is mapped a relatively high resolution texture completely unrelated to the actual surface of Apophis. The spherical shape is used in our analysis in order to provide an easy access to the ground truth of landmark positions, knowing the true camera poses. In addition, the lighting conditions representative of a terminator orbit have been taken into account (only one asteroid hemisphere is lit).

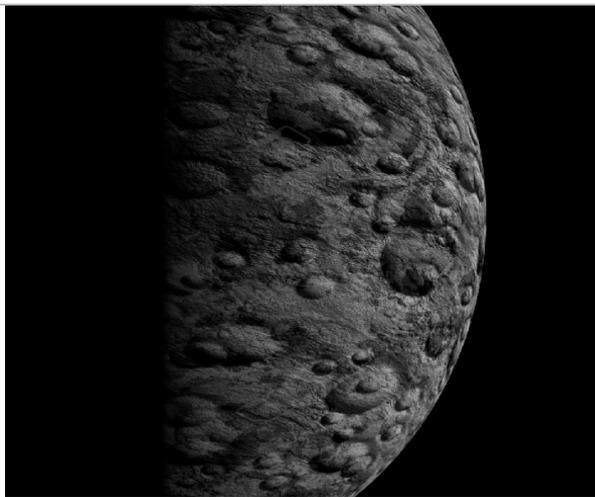


Figure 6.1: Simulation of the asteroid aspect

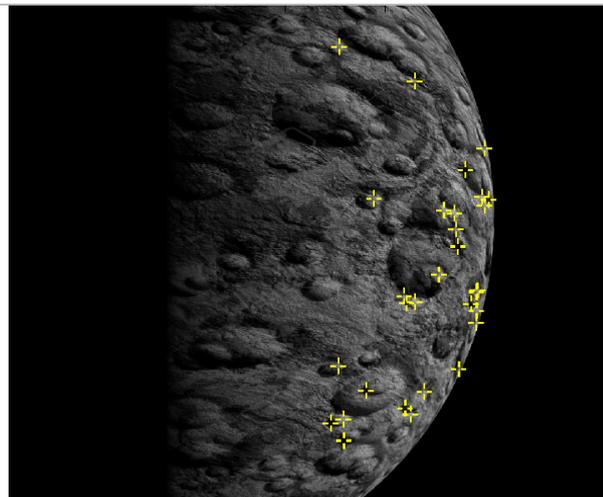


Figure 6.2: Landmarks detected on the asteroid surface (30 best scores)

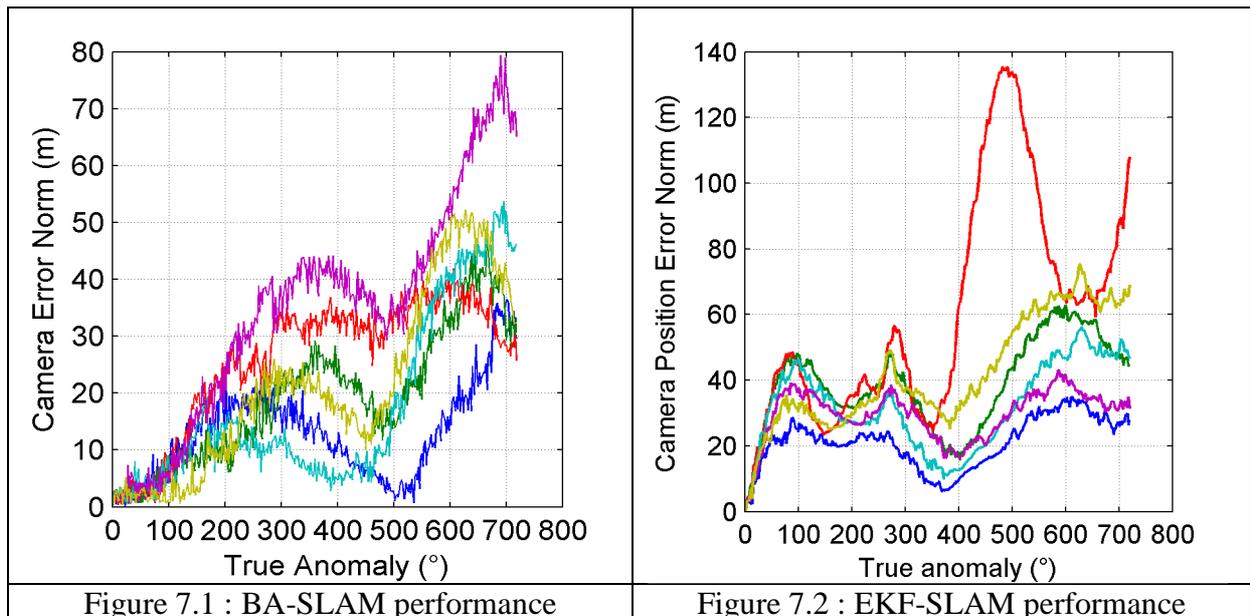
Image processing: The implementation is based on the OpenCV SURF algorithm that has been integrated within a Matlab S_Function. Figure 6.2 shows an example of features that this algorithm can detect in images of the asteroid surface. Here only 30 features with the highest

strength response are shown for clarity, but over 1500 were actually detected. We can observe that the strongest features are naturally attracted to high-contrast regions of the surface that are located in the region that benefits from the best lit conditions. The repeatability of this detector has been evaluated on the simulated images and the ratio is around 90-95% considering 5° viewpoint changes.

As regards feature matching between successive views, it is performed in two steps. The first one relies on the SURF matching functionality that computes the Euclidian distance between descriptors and returns the pairs of features with the shortest distance. To filter the incorrect matches, we apply afterwards a matrix based RANSAC criteria taking advantage of the images epipolar geometry. This algorithm implemented from the RVCTools toolbox [21], allows producing the largest set of matches that agree on the same transformation between two images. This filtering is highly efficient to remove outliers but this benefit comes at the expense of a certain amount of inliers that are also rejected.

Even though this method is rather computationally expensive, it has been preferred to the popular active search approach that relies on the position uncertainty knowledge and is thus applicable only in combination with the EKF-SLAM algorithm. It must be noted that the computational cost of the feature detection and matching algorithms averages 3 to 4 s in our implementation and outweighs by far the cost of the estimation part.

Vision layer coupling with the estimation algorithms: The feature detection and matching algorithms have been coupled to the BA-SLAM and EKF-SLAM to assess their performance in more realistic conditions. For this experiment, the implementation is not a true integration since measurements are fed to the estimators but the vision layer does not benefit from the estimated state. The vehicle trajectory corresponds to 2 revolutions on a terminator orbit, images are acquired every 5° anomaly increment. The asteroid being fixed, a total of 72 images allow creating a sequence that can simulate a perpetual and continuous motion. Using the same set of measurement data, multiple simulations can be run by selecting different initial poses in the sequence. Figure 7.1 and 7.2 show the behavior of both estimators for 6 different runs corresponding to initial poses separated by 60° anomaly increments.



Even though both estimators remain functional, some performance degradation with respect to the analysis in section is observed with a noticeable increase in the mean error and standard

deviation characteristics. The impact is the most significant on the EKF-SLAM behavior that shows a quasi divergence on one of the simulation cases. This degradation can be explained by an irregular landmark distribution covering a smaller part of the field of view, the larger amplitude of the measurement noise and the presence of a few remaining outliers. The drift magnitude that is observed seems acceptable at this stage since it is assumed that some navigation update from the ground can be performed every two revolutions (roughly 2 days).

6. Conclusion

This study has focused on the feasibility of using monocular vision-based estimation techniques in the context of navigation and mapping while operating in close proximity of an asteroid. The main goals of the study were to implement, test and compare the performance of two distinct SLAM navigation techniques: a sequential Bundle Adjustment algorithm (BA-SLAM) that performs optimization over multiple views and a classical Extended Kalman Filter (EKF-SLAM) which is more adapted to real-time execution. First, a parametric study was performed to compare the methods in presence of a virtual visual layer. The preliminary results showed that BA-SLAM is functional starting with 3 views only, yields better performance than EKF-SLAM in similar conditions for a computational cost only a few times higher. In absence of range measurement or external position updates, both estimators are subject to drift during the orbital motion and the error amplitude can be rather dispersed after a few revolutions. As expected, tracking a larger number of landmarks enabled to reduce the drift and the dispersion and some improvement was also obtained to some extent by increasing the number of views. The work was completed with the coupling of both estimators to a vision layer processing simulated asteroid images and this allowed to confirm the performance behavior of both methods in more representative conditions. This showed also that the dominant computational burden is definitely on the image processing side and makes the debate about the estimation execution cost less relevant.

These results are promising but they were obtained under a set of simplifying assumptions that need to be revisited to broaden and consolidate the evaluation analysis. Several important aspects remain to be characterized: the sensitivity to initial state errors, measurement outliers, residual uncertainty on the asteroid motion, the influence of more realistic vehicle dynamics and landmark distribution due to a complex asteroid shape and texture. From the BA-SLAM perspective, the absence of explicit uncertainty knowledge constitutes a limitation that needs to be overcome to improve the efficiency of feature matching. Robustness being a central issue, a continuation of this study will also address two topics: the value of sparse distance measurements to improve position observability, the re-detection of landmarks already observed in the past. Another path that remains to be explored is the applicability of a hybrid algorithm, merging the best elements of EKF and BA-SLAM together. As a subsequent step, it would also be essential to evaluate the hardware requirements of one of the considered navigation methods, which could give an insight on how far in the future autonomous, embedded and vision-based navigation could be envisioned.

7. References

- [1] R. Gaskell, et al. (2006), "Global Topography of Asteroid 25143 Itokawa," 37th LPSC, Houston.
- [2] S. Ulamec, J. Biele, "Surface elements and landing strategies for small bodies missions – Philae and beyond", *Advances in Space Research*, Volume 44, Issue 7, 1 October 2009
- [3] Owen, W. M., Wang, T. C., Harch, A., Bell, M., and Peterson, C., "NEAR Optical Navigation at Eros," Paper AAS 01-376, *AAS/AIAA Astrodynamics Specialists Conf.*, Quebec City, Quebec, Canada (30 Jul–2 Aug 2001).

- [4] N. Mastrodemos et. al, "Optical Navigation for Dawn at VESTA", *American Astronautics Society - AAS* 11-222.
- [5] P. Michel, J.Y. Prado et al., "Probing the interior of asteroid Apophis: a unique opportunity in 2029", *Proceedings of the International Astronomical Union / Volume 10 / Highlights H16 / August 2012*, pp 481-482
- [6] D. Conway, B. Macomber, K. Cavalieri, J. Junkins, "Vision Based relative navigation filter for asteroid rendezvous".
- [7] C. Cocaud, T. Kubota, "Autonomous navigation near asteroids based on visual SLAM" . *23rd International Symposium on Space Flight Dynamics*, Pasadena (Oct 2012)
- [8] T. Lemaire, S. Lacroix, and J. Sola. A practical 3d bearing-only slam algorithm. *In Intelligent Robots and Systems, (IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 2449-2454. IEEE, 2005.
- [9] Georg Klein and David Murray. Parallel tracking and mapping for small AR workspaces. *In Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 225-234. IEEE, 2007.
- [10] J. Sola, "Consistency of the monocular EKF-SLAM algorithm for three different landmark parametrizations". *In Robotics and Automation (ICRA), 2010 IEEE Intern. Conf. on*, pages 3513-3518. IEEE, 2010.
- [11] J. Neira and J.D. Tardos. Data association in stochastic mapping using the joint compatibility test. *Robotics and Automation, IEEE Trans. on*, 17(6):890-897, Dec 2001.
- [12] J. Civera, O. G. Grasa, A. J. Davidson, and J.M.M Montiel. "1-point ransac for extended kalman filtering : Application to real-time structure from motion and visual odometry". *Journal of Field Robotics*, 27(5):609-631, September 2010.
- [13] Strasdat. H., Monteil J.M.M, Davison A.J., Visual SLAM : Why filter ? *In Robotics and Automation (ICRA), 2010 IEEE International Conference on*,
- [14] C. Engels, H. Stewenius, and D. Nister. Bundle adjustment rules. *Photogrammetric computer vision*, 2, 2006.
- [15] D. G. Lowe. Object recognition from local scale-invariant features. *In Proceedings of International Conference on Computer Vision*, pages 1150–1157, 1999.
- [16] T. Tuytelaars and K. Mikolajczyk. Local Invariant Feature Detectors: A Survey. *Foundations and Trends in Computer Graphics and Vision* Vol. 3, No. 3 (2007) 177–280
- [17] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. 2011. "ORB: An efficient alternative to SIFT or SURF." *IEEE Computer Vision (ICCV)*.
- [18] Lowe. D. Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision*, 2004
- [19] Kobayashi. T., Simon. D., Litt. J., "Application of a Constant Gain Extended Kalman Filter for In-Flight Estimation of Aircraft Engine Performance Parameters"
- [20] [https://en.wikipedia.org/wiki/Blender_\(software\)](https://en.wikipedia.org/wiki/Blender_(software))
- [21] P.I. Corke, "Robotics, Vision & Control", Springer 2011, ISBN 978-3-642-20143-1.