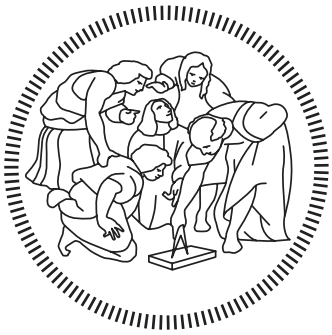
School of Industrial and Information Engineering
Department of Aerospace Science and Technology
Master of Science in Space Engineering



POLITECNICO
MILANO 1863

MULTISPECTRAL VISION-BASED RELATIVE NAVIGATION TO ENHANCE SPACE PROXIMITY OPERATIONS

Advisor:
Prof. Michèle Lavagna
Co-Advisors:
Gaia Letizia Civardi
Margherita Piccinin

Candidate:
Massimiliano Bussolino
964727

Academic Year 2021-2022

Abstract

CLOSE-proximity operations play a crucial role in new classes of missions, such as Active Debris Removal. Approaching a non-cooperative target, the elevated risk of collision and the scarce reliance on ground intervention dictate the need for autonomous onboard pose (attitude and position) estimation for the chaser spacecraft. Monocular cameras operating in the visible (VIS) spectrum have been widely investigated and in-flight tested for navigation applications because of their adequate balance between the quality of the provided data and the limited mass and power consumption. However, the strong dependency of the results on the target's illumination condition strongly reduces this sensor's application range. A frequently investigated solution to mitigate this problem is the implementation of a monocular thermal-infrared (TIR) camera concurrently with the visible one. Thermal sensors provide lower-quality data than visible sensors but they are less dependent on illumination conditions. This thesis proposes a novel visual navigation pipeline to perform pose tracking with measurements obtained from a VIS and a TIR monocular camera. The navigation algorithm exploits the point features extracted from the images to perform model-based pose estimation with respect to a known target. The point features detected in the multispectral images are fused with a tightly coupled approach through an Extended Kalman Filter, whose output is the refined pose of the chaser. Such innovative pipeline will be presented in detail in this thesis, along with tests performed with synthetic images. The proposed navigation chain successfully performs the pose estimation: results will be presented, assessing the contribution of combining a TIR and a VIS monocular camera, studying the performance of the fused multispectral information and characterizing the performance of TIR-only navigation in low-illumination conditions. This thesis is intended to contribute to the assessment of the performance and robustness of multispectral visual navigation and developing pose estimation schemes which operate in demanding environments like Active Debris Removal.

Keywords: Proximity operations, Visual-based relative navigation, Non-linear estimation, Multispectral data fusion

Sommario in italiano

Le operazioni di prossimità ravvicinata ricoprono un ruolo cruciale nelle nuove classi di missioni, come la Rimozione Attiva di Detriti (ADR). Avvicinandosi ad un target non cooperativo, l'elevato rischio di collisione e la scarsa possibilità di affidarsi ad un intervento da terra richiedono la stima autonoma della posa (assetto e posizione) dell' inseguitore. Le fotocamere monoculari nello spettro visibile sono state ampiamente studiate e testate per applicazioni di navigazione visto il loro bilanciato rapporto tra la qualità dei dati che forniscono e le ridotte massa e potenza richieste. Tuttavia, la forte dipendenza dei dati forniti dalla condizione di illuminazione del target riduce fortemente le possibilità di applicazione del sensore. Una soluzione frequentemente proposta per mitigare questo problema è l'implementazione di una fotocamera monoculare termica insieme a quella visibile. I sensori termici forniscono dati di minor qualità rispetto a quelli visibili, ma sono meno dipendenti dalle condizioni di illuminazione. Questa tesi propone una catena di navigazione innovativa per effettuare il rilevamento della posa tramite misurazioni ottenute da una camera visibile e una termica. L'algoritmo di navigazione usa i punti d'interesse estratti dalle immagini per effettuare la stima della posa con un approccio model-based rispetto ad un obiettivo noto. I punti d'interesse identificati nelle immagini multispettrali sono fusi con un approccio strettamente accoppiato attraverso un Filtro di Kalman Esteso (EKF), il cui output è la posa corretta dell' inseguitore. La tesi presenta il suddetto algoritmo, che effettua con successo la stima della posa, insieme ai test effettuati su immagini sintetiche. Si valuta il contributo della combinazione delle fotocamere visibili e termiche, analizzando inoltre la possibile applicazione della navigazione solo termica in scarse condizioni di illuminazione. La tesi vuole contribuire alla comprensione delle prestazioni e della robustezza della navigazione ottica multispettrale e a sviluppare algoritmi per la stima della posa che possano operare in ambienti difficili come quello della Rimozione Attiva di Detriti.

Parole chiave: Operazioni di prossimità, Navigazione relativa ottica, Stima non lineare, Fusione di dati multispettrali

Ringraziamenti

Vorrei ringraziare la Professoressa Michèle Lavagna per avermi dato l'opportunità di lavorare a questa tesi e per avermi convolto in progetti ed esperienze fondamentali per la definizione del mio futuro. Inoltre, ci tengo a ringraziare Gaia e Margherita per la disponibilità e il supporto di questi mesi, permettendomi di svolgere il lavoro sempre con sicurezza e serenità.

Un ringraziamento sincero va a mia madre e a mio padre, la cui cieca fiducia nonostante la mia riservatezza è da sempre la base fondante dei risultatiti ottenuti. A mia sorella che, quando con l'esempio e quando con la motivazione, mi ha sempre spinto, e sempre mi spingerà, a un di più che mai avrei raggiunto da solo.
A Claudia, per avermi sempre sostenuto, per riordinarmi i pensieri quando tutto sembra confuso.

Grazie infine a tutti coloro che hanno condiviso con me questo percorso, rendendo memorabili i momenti piacevoli e leggeri quelli difficili; per esserci sempre stimolati senza mai competere.

Contents

List of Figures

List of Tables

List of Algorithms

List of Acronyms

AAM Active Appearance Model.

ADR Active Debris Removal.

AKE Absolute Knowledge Error.

AVANTI Vision Approach Navigation and Target Identification.

CLAHE Contrast Limited Adaptive Histogram Equalization.

CNN Convolutional Neural Network.

CPO Close Proximity Operations.

CV Computer Vision.

CW Clohessy - Wiltshire.

DARPA Defense Advanced Research Projects Agency.

DCM Direct Cosine Matrix.

DLR German Aerospace Center.

DLT Direct Linear Transform.

ECI Earth Centered Inertial.

ECSS European Cooperation for Space Standardization.

EKF Extended Kalman Filter.

ESA European Space Agency.

FF Formation Flying.

FOV Field of View.

GNC Guidance Navigation and Control.

GNSS Global Navigation Satellite System.

IADC InterAgency Debris Coordination Committee.

IDO Intact Derelict Object.

IMU Inertial Measurement Unit.

ISO International Organization for Standardization.

JAXA Japan Aerospace Exploration Agency.

KF Kalman Filter.

LEO Low Earth Orbit.

LiDAR Light Detection And Ranging.

LK Lucas-Kanade.

LVLH Local Vertical local Horizon.

MEKF Multiplicative Extended Kalman Filter.

MKE Mean Knowledge Error.

MRP Modified Rodriguez Parameters.

OOS On Orbit Servicing.

PCA Principal Component Analysis.

PnP Perspective-n-Point.

RANSAC RANdom SAmple Consensus.

RKE Relative Knowledge Error.

SLAM Simultaneous Localization and Mapping.

STL STereolithograpfy File.

STM State Transition Matrix.

TIR Thermal-InfraRed.

TIRI Thermal InfraRed Imager.

TM Template Matching.

UKF Unscented Kalman Filter.

VESPA VEga Secondary Payload Adapter.

VIS VISible.

1

Introduction

Vision based relative navigation is becoming increasingly important to enable the development of next-generation space missions, such as the exploration of natural celestial bodies, On Orbit Servicing (OOS), Formation Flying (FF) or Active Debris Removal (ADR). In this context, the necessity of real-time guidance and decision making in a dynamical environment makes onboard relative attitude and position estimation paramount for the successful completion of the mission objectives.

Within the previously mentioned mission scenarios, ADR has recently gained significant interest because of the growing concern about space situational awareness. Dealing with inherently non-cooperative targets, ADR mission shall be equipped with vision-based sensors able to provide a high-frequency relative pose estimation so to avoid undesired collisions with the target.

In this context, monocular cameras in the visible spectrum present a very mass and power-effective solution, although presenting important operational limitations. Their strong dependency on the illumination conditions of the target limits their range of applications in case of low illumination conditions or eclipse. A promising solution to enhance the reliability of monocular cameras is to use collaboratively two sensors working in the visible and thermal spectrum. Thermal imaging sensors measure the thermal radiance of the target, which is influenced by the illumination coming from the Sun, but it is not subject to shadowing or reflections. However, the overall quality provided by thermal imaging sensors is lower because the images produced are subject to a higher blur and a lower resolution.

In this framework, the thesis proposes a novel navigation pipeline that exploits an Extended Kalman Filter to perform feature-level adaptive sensor fusion between the two cameras. The stand-alone vision-based navigation chain is tested on synthetically generated images to assess its performances, evaluating as well the capability to perform TIR-only navigation in low illumination conditions.

The present chapter further details the motivation for the thesis work and presents the literature related to the investigated subject. Finally, the work's intended contribution is presented with a general overview and the thesis outline.

1.1 Context & motivation

1.1.1 Active Debris Removal

Since the beginning of spaceflight, the amount of artificial objects in orbit has increased steadily. In particular, non-cooperative objects, better known as space debris, threaten satellites operating in certain orbital families such as Low Earth Orbits. Awareness of this problem led to the publication in 2002 of the IADC Space Debris Mitigation Guidelines [61] by the InterAgency Debris Coordination Committee, superseded by the 2011 ISO standard 24113 [26] on debris mitigation requirements. These standards have been adopted by the European Committee for Space Standardization (ECSS) and apply to all projects developed by the European Space Agency (ESA).
However, the effect that such recommendations have had is limited. Firstly, entities operating space missions are not required to follow them, as they are not applicable regulations. Furthermore, even if these were strictly followed, the environment is unstable [40], i.e., the number of debris will increase even if all space activities are interrupted because of consecutive impacts of existing objects. An identified solution to limit the growth of the number of space debris is to actively deorbit human-made objects with ADR missions. According to [33], removing five Intact Derelict Objects (IDOs) per year would stabilize the situation in LEO. IDOs comprises mainly inoperative satellites, such as the European debris ENVISAT, launcher stages and payload adapters.
ADR missions aim at interacting with space objects to reduce their orbital lifetime. In this context, the active spacecraft is referred to as chaser, while the spacecraft to be decommissioned is the target. The mission profile of an ADR mission can be subdivided into different operational phases, presented in Figure 1.1. After being launched, the chaser phases with the target orbit reaching an hold point, relying on absolute navigation. Subsequently, during the inspection phase, the chaser collects information about the target with onboard sensors to plan the subsequent phases. The duration of this phase and the distance maintained from the target depends on the prior knowledge about the target, and the characteristics of the sensors of the chaser. For the inspection, it is necessary to rely also on relative navigation, as the shorter the chaser-target distance, the higher the risk of collision. The information collected during the inspection phase is confirmed and improved during the fly-around phase, further approaching the target. Once gathered the required information the chaser performs the final approach and capture in order to capture the target. Finally, the target is safely disposed in the stabilization and de-orbiting phase.
The first mission to perform this task will be the ClearSpace-1 mission in 2025 [8], commis-

sioned by ESA. The mission aims to deorbit VEga Secondary Payload Adapter (VESPA).

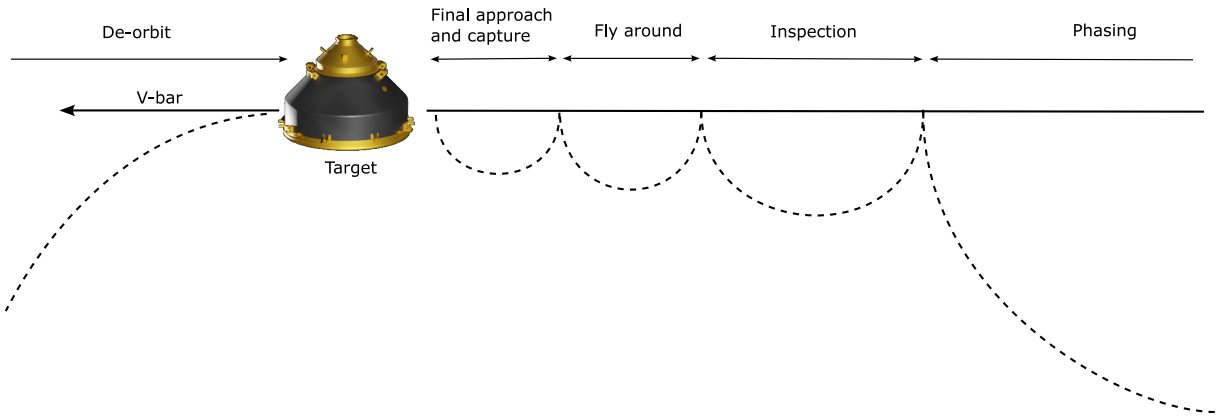


Figure 1.1: Simplified schematic of the typical phases of an ADR mission

1.1.2 Autonomous visual navigation

For applications including Close-Proximity Operations (CPO) and Active Debris Removal, it is paramount to have a high degree of onboard autonomy. To avoid the risk of an undesired contact with the target, the chaser shall be able to operate autonomously to be robust against communication delays or unavailability. In this framework, the necessity to have continuous relative position and attitude estimation with respect to the target is evident.

Unlike Formation Flying or In-Orbit Assembly scenarios, in the case of navigation around space debris, the target is always non-cooperative. Therefore, visual navigation must be used to estimate the relative position and attitude, using optical sensors such as Light Detection and Ranging (LiDAR), stereo cameras, or monocular cameras. For the selection of the navigation sensor, its precision and range of applicability are paramount. However, the trade-off shall consider parameters like mass, power consumption, the processing power required, and cost. Monocular cameras represent a more mass and power-effective solution than LiDAR while granting a more comprehensive operational range than stereo cameras [43, 45].

The Orbital Express carried out the first in flight-demonstration of autonomous docking to a non-cooperative spacecraft in 2007 [47], commissioned by DARPA, using optical images and a system of corner-cube reflectors. Additional flight heritage is given by the PRISMA mission in 2010 [28], providing a more relevant example of the usage of monocular cameras. The mission was composed of two spacecrafts: Mango and Tango. The Vision Based Sensor mounted on Mango comprised two monocular cameras, one for far range

and one for close range operations, working in the visible spectrum. Within the multiple mission objectives, Mango performed an autonomous vision-based rendezvous with Tango acting like a non-cooperative target. Once this was completed, an additional experiment, IRIDES, was added to test the visual navigation pipeline to rendezvous and inspect the uncooperative Picard spacecraft. However, the rendezvous could not be tested because of the delta-v depletion before the phasing with Picard [21]. Finally, to consolidate the results of the PRISMA mission, the German Aerospace Center (DLR) launched the Autonomous Vision Approach Navigation and Target Identification (AVANTI) experiment onboard the BRIOS spacecraft [22], aimed at demonstrating autonomous rendezvous from 10 km to 1 km towards an uncooperative spacecraft making use of visual angles-only measurements. To achieve this result, the same far-range navigation camera as PRISMA was mounted on BRIOS.

Despite the significant progress made in the last decades and the flight heritage given by past missions, navigation technologies for autonomous relative navigation about uncooperative targets still require further development to de-risk them and increase their robustness. Moreover, the development of more reliable navigation solutions for the ADR application would have a positive effect on other mission scenarios. The increased interest in missions involving OOS and FF requires improving the current state of the art of navigation technologies. A planned mission aimed at contributing to this aspect is the ESA mission e.Inspector, which includes within its mission objectives to test optical technologies for navigation about an uncooperative target. Within the proposed architecture to be mounted onboard e.Inspector, currently in phase B of its development, it is proposed a combination of visible and thermal monocular cameras. The addition of the thermal camera serves to increase the range of applications of the visible one to different illumination conditions.

1.1.3 Thermal imaging

Because of their wide range of applications and long history of use, cameras working in the visible spectrum (VIS) provide high-quality data with low power consumption. This is not the case for thermal (TIR) cameras. TIR cameras based on cooled microbolometers, working in the Mid-Wave (3 µm - 8 µm) to Long-Wave (8 µm - 15 µm) infrared, are not suitable for space applications due to the power required and the complexity of the cryogenic cooling of the bolometer. The state of the art for space scenarios consists of uncooled sensors working only in the Long-Wave infrared, providing sufficient sensitivity without the issues of cooling the sensor [10]. The flight heritage of this sensor is relatively limited, as it was tested only on the LIRIS experiment onboard the ATV5 Mission [11]

and later on the Raven ISS Hosted Payload [23].
Additional heritage is provided in the context of asteroid exploration. The Hayabusa 2 mission [41] acquired close range thermal images of the Ryugu asteroid in the Long Wave InfraRed spectrum. The thermal imaging was not performed for the purpose of relative navigation, although it was noted that in the thermal spectrum it was possible to clearly identify visual markers placed on the surface as cold spots, highlighting the possibility of multispectral navigation. This technology will be tested for asteroid navigation in the ESA mission Hera [38], mounting the TIRI (Thermal InfraRed Imager) payload provided by JAXA (Japan Aerospace Exploration Agency).
Compared to VIS cameras, TIR sensors are less sensitive to different illumination conditions because they depend on the thermal profile of the target and its emitted radiation. Thus, the thermal cameras could be employed also when the target is not illuminated if its temperature is within the sensor's sensibility ranges. However, the image resolution is lower than VIS cameras, with a higher blur that reduces its performance for relative navigation. Because of the limitations of both visible and thermal imaging sensors, combining the two represents a promising solution for increasing the robustness of relative optical navigation.

1.2 Literature review

This section presents the current state of the art in multispectral visual navigation. The literature is reviewed with a top-down approach, beginning with an overview of monocular pose estimation methods, followed by a discussion of common filtering approaches for relative navigation, and concluding with an examination of the latest developments in multispectral data fusion for navigation applications.

1.2.1 Monocular pose estimation

Monocular pose estimation is the process estimating a target spacecraft's relative attitude and position with respect to a chaser spacecraft, using the measurements output of a single monocular camera or multiple cameras fused together.
The taxonomy of pose estimation methods divides them based on how the image information is exploited. In appearance-based approaches, also known as direct methods, the raw visual measurement in terms of pixels is used [39]. Algorithms such as the Active Appearance Model (AAM) or the Principal Component Analysis (PCA) are used to match the object detected in the image to a stored database of the target images. This approach, although often implemented in robotics and surveillance applications, is less suitable for

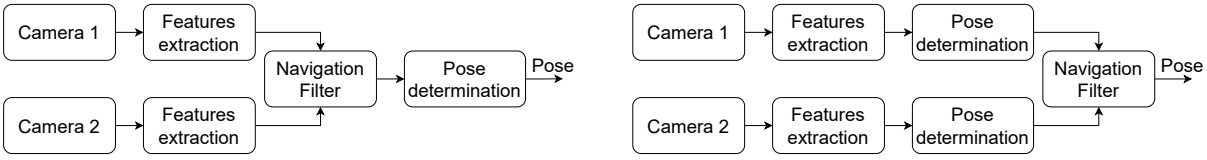
space applications because of its scarce robustness to noise in the images or elements in the background, which are common elements of space imagery [10, 42].
More suitable approaches for space applications are the feature based-methods, which exploit relevant features of the image, such as corners or edges, to reconstruct the target's pose. These methods apply to navigation about both known and unknown targets. If the shape of the target is not known, a common solution is to reconstruct it with the available measurements. This is commonly accomplished with Simultaneous Localization and Mapping (SLAM) [19], a real-time process where the chaser computes at the same time its location (Localization) and a map of the target (Mapping). As the navigation about unknown objects is outside the thesis work, the reader is referred to [42] for a broad overview of uncooperative pose determination about unknown targets.
In the case of navigation about a known target whose 3D model is available a priori, or if the model has been reconstructed via SLAM, this information can be used to enhance the pose estimation process. In Perspective-n-Point (PnP) solver approaches, the features extracted from the image are matched to the wire-frame model of the target to retrieve the pose by solving the PnP problem. Because of the lack of fiducial markers on the target, the matching problem can result challenging. If corners, also known as point features, are extracted from the image, a solution is represented by the SoftPosit algorithm [17]. The algorithm provides a joint solution to the two aspects using the soft-assign technique to find correspondences and the POSIT (Pose from Orthography and Scaling with Iterations) to compute the relative pose. This approach has been successfully tested for spacecraft relative navigation in [55] and [54], although noticing a decrease in accuracy when given an initial guess far from the solution. An alternative is to match with the RANdom SAmple Consensus (RANSAC) algorithm [20] then use a Perspective-n-Point solver. A commonly used algorithm to solve the PnP problem is the Efficient Perspective-n-Point (EPnP) [31], as it offers a non-iterative solution. An example of this implementation can be found in [48], successfully tested on synthetic images of the ENVISAT inoperative satellite. Alternatively, other elements can be identified as corner features, such as edges, as investigated by [53], or more complex features as circles or ellipses [34]. Edge features are found to be more robust to different illumination conditions, although less reliable when the target is only partially visible. Moreover, if more complex shapes such as circles are used, they apply to spacecraft with a specific geometry only, limiting their applicability.
Another approach used to estimate the pose of a known target exploiting features is Template Matching (TM). This method requires the offline generation of a database of template images of the target by sampling the six degrees of freedom (position and attitude). The most similar image to the one captured by the camera is identified, and the pose is automatically retrieved from the database. Even if the online computational

effort might be reduced with respect to PnP solvers, the lack of robustness to illumination conditions makes it less appealing for space applications.
Because of their fast development in recent years, Convolutional Neural Networks (CNNs) are of increasing interest for monocular pose estimation. Unlike previously described methods, CNN-based methods do not distinguish from a feature extraction and pose estimation phase but rather an offline training and online test. Their application is expected to improve the robustness with respect to existing methods and reduce the computational burden. In [52], a CNN-based navigation algorithm is validated on a database of images comparable to the actual space images taken by the PRISMA Mission, obtaining better results than classical methods. Learning methods are currently limited by the need for large databases of authentic space imagery or synthetically generated images necessary for their training [10].
In recent years, the possibility to perform monocular navigation with a thermal camera has been investigated. In [64] the beneficial effects of using thermal imagery are illustrated, also studying the images acquired by the LIRIS demonstrator, showing that a SLAM-based approach is suitable for thermal cameras. This result is also confirmed in the context of asteroid navigation in [12, 45, 46]. Contrary to the visible spectrum, there is a lack of literature regarding the generation of realistic space thermal imagery, necessary for the validation of the proposed architectures. The recent works of [13, 65] fill this gap by providing new insights and methods for generating realistic space thermal imagery.

1.2.2 Visual-based navigation filters

The pose estimation schemes introduced to this point can provide an estimate of the relative position and attitude of the target without any a priori information. When an initial estimate is provided from a lost-in-space condition, it is commonly referred to as pose acquisition or pose initialization. Once an initial estimate is provided, pose tracking can be performed, using the previous pose estimate for each new image acquired as the initial condition for the pose acquisition problem. Pose initialization schemes are unsuited to work at high frequencies because of their high computational time. Therefore, a visual navigation filter shall be used to process the camera measurements to provide the updated pose at high frequency [10].
In relative navigation, two different architectures are distinguished based on which level the filter processes the measurements. In tightly coupled filters, the extracted features are directly processed by the filter, reconstructing the pose with the refined output of the filter. Alternatively, in loosely coupled filters the pose is already determined prior to the filter. A schematic of the two architectures is reported in Figure 1.2, considering a

navigation architecture using two cameras. The same concept applies to other types of sensors, such as IMU or GNSS [29].



(a) Tightly coupled **(b)** Loosely coupled

Figure 1.2: Tightly and loosely coupled navigation filter architecture using two cameras

Loosely coupled are usually preferred for applications where the target is fully known, as the computational time of tightly coupled approaches depends on the number of features detected and they lack robustness to fast relative dynamics [43]. However, if the inertia properties of the target are unknown, tightly coupled filters can retrieve them [60].
A very widely used filter in relative navigation applications is the Extended Kalman Filter (EKF) [23, 43, 53] because of their balanced ratio between low complexity and sound performances. As relative dynamics are intrinsically non-linear, the EKF exploits the linearization of the dynamics. In the case of highly non-linear dynamics, non-linear filters are implemented. In the works [60, 67], the Unscented Kalman Filter is exploited to improve the filter's accuracy, but with the degradation of the computational efficiency. In [44], several non-linear filtering techniques such as Minimum Energy Filter, 2nd Order Energy Filter, and Attitude Observer were investigated, uncovering that they outperform linear filters.

1.2.3 Multispectral data fusion

Multispectral navigation is an emerging area of interest for space applications, yet the literature on this topic is still limited. Data fusion can be performed at different levels. The lowest level is to perform image fusion (image-level data fusion), creating from two images in different spectra a composite image with the best characteristics of the two. An extensive survey of methods for image fusion is reported in [56]. In [14], different methods are applied to the fusion of space imagery performing a comparative assessment. These findings were applied in [6], where fused images were used to perform pose initialization, demonstrating that fused images can be a reliable source of measurements for navigation algorithms.
If the information fusion is performed at a higher level, then the problem can be approached with classical sensor fusion techniques. Odometry fusion techniques are divided into optimization-based, and filter-based techniques [63]. In the former, the state is estimated after processing the measurements acquired into an optimization problem. These

methods are more accurate than filter-based methods, although presenting a higher level of complexity [27].
In filter-based approaches, the sensor information is fused inside the navigation filter, as described in Section 1.2.2. This methodology has been investigated for multispectral sensor fusion in the context of the ESA study "Image Recognition and Processing for Navigation (IRPN)" [57]. The interest in this approach was also expressed in [23], fusing VIS, TIR, and LiDAR measurements employing a loosely-coupled Kalman Filter. To the best of the author knowledge, a tightly coupled approach for multispectral data fusion has not been investigated for space applications.
Further research has been performed in the context of asteroid navigation. In particular, [12] uses a loosely coupled approach to fuse multispectral data, while [45] investigates both a multi-modal approach and image fusion. Both studies focus on a SLAM algorithm, as the map of the asteroid is not known a priori.
The lack of literature in the field of multispectral data fusion applied to relative navigation highlights the need for further research in this area, aiming at paving the road for robust navigation pipelines for harsh environments as the one of Active Debris Removal.

1.3 Research objectives and overview

The paragraph highlights the thesis scope and its logical flow. It presents the research objectives and formulates them as research questions. The thesis overview and the document structure is presented.

1.3.1 Objectives

The core motivation of the thesis is the enhancement of visual relative navigation techniques about uncooperative targets whose shape is known, given the context explained in Section 1.1. In particular, the principal aim of the thesis is:

> *to investigate the potential in exploitating of visible and thermal imaging measurements fusion for enhancing relative navigation about known uncooperative space targets.*

The thesis aims at providing a deeper understanding of the contribution of multispectral imaging to the estimation of the relative attitude and center of mass position. Specifically, it investigates how sensor fusion expands the range of applicability and increases the performances of monocular visible navigation by combining data from multiple spectra.

Moreover, the effectiveness of thermal navigation as a standalone solution is presented, studying the possibility of substituting the visible counterpart when the output of the latter is compromised. In this context, the effort is focused on assessing most suitable strategy between privileging either visible or thermal measurements depending on environmental conditions or going straight forward with data fusion always, no matter of external conditions. The work proposes a solid basis for a comparison with different data fusion techniques, providing further understanding about at which level the multispectral data shall be fused to obtain the most proper results.
Additionally, this work provides valuable insights into best practices and potential pitfalls whenever designing simulation frameworks for testing visual navigation algorithms.

1.3.2 Research questions

The thesis objective is then articulated in the following research questions:

1. can multispectral imaging improve relative navigation about a known non-cooperative target? What is the contribution of thermal imaging?
2. Which is the most effective visible and thermal imaging measurements management to improve relative navigation?
3. Can thermal imaging sensors provide a standalone solution to the relative navigation problem? What are the most prominent limitations of *thermal-only* navigation?
4. Upon what criteria VIS-IR measurements fusion is preferred against treating them as split inputs? Is there an overall winner among the tightly, loosely coupled and image-fusion based approaches?

1.3.3 Thesis overview

To develop the aforementioned objective, this thesis proposes a novel multispectral navigation pipeline for pose tracking with measurements obtained from a VIS and TIR monocular cameras. The algorithm exploits the information extracted from the images to perform model-based pose estimation with respect to the target. The sensor fusion and the pose refinement are performed with an Extended Kalman Filter. The multispectral information is fused at feature level with a tightly coupled approach. In particular, the features tracked on the VIS and TIR images represent the measurements fed into the Kalman Filter, while the pseudo-measurements are computed by projecting the model landmarks onto the image plane. Such fusion technique enables avoidance of the execution of pose reconstruction routines such as the PnP, without the need to expand the filter's state

vector, as commonly performed in tightly coupled approaches. In the proposed solution, an online adaptation of the filter's characteristic matrices is employed to tune the contribution of the different features based on the data quality.
Although the proposed algorithm's primary goal is to investigate the benefits of combining multispectral imaging sensors, this work is intended to be the basis for a future onboard implementation of the algorithm. For this reason, special attention is given to the computational time required and the complexity of the steps of which the navigation chain is composed.
The proposed navigation chain is validated exploiting synthetic images of a space debris. The results are presented and discussed to reach the thesis objective and answer to the research questions.
The thesis is structured as follows:

- **Chapter 2** provides a brief explanation of the principles at the basis of the thesis, ranging from aspects of computer vision to preliminary aspects of non-linear estimation methods used.
- **Chapter 3** presents the proposed visual navigation pipeline, introducing both the filter structure and the IP algorithms used.
- **Chapter 4** describes the simulation framework used to validate the proposed solution to the estimation problem.
- **Chapter 5** presents the simulations' outcome, providing a critical analysis of the obtained results.
- **Chapter 6** summarizes the key findings, identifying the final conclusions. Ideas for future works are also reported.

2

Background

This chapter provides a brief explanation of the principles and algorithms at the basis of the thesis.
As the visual navigation pipeline strongly relies on computer vision, some basic aspects of this subject are introduced. Initially, the pinhole camera model is presented, describing the relation between world points and their projection on the image plane. Subsequently, projective transformations between two sets of points are introduced, focusing on the Direct Linear Transform (DLT) algorithm to estimate the homography associated with the transformation. The concept of key points, or point features, is presented. The description presents some widely used methods for detection, description, and tracking, setting the framework for the analysis and trade-offs performed during the thesis development.
Finally, a brief discussion on non-linear estimation methods is presented, focusing on the multiplicative extended formulation of the Klaman Filter (KF) for navigation applications.

2.1 Computer Vision

Computer Vision (CV) is a multidisciplinary subject that enables computers to derive meaningful information from digital images. This section describes some preliminary aspects on which the thesis work is built upon.

2.1.1 Camera model

A camera model represents the mathematical relationship between 3D points and their respective 2D projection on the image plane. In particular the pinhole camera model is a simple but faithfully representative model of optical navigation cameras. The camera model representation is reported in Figure 2.1.

Under the pinhole camera model, real-word points are projected onto the image plane (or focal plane) $Z = f$, resulting in a transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$. More specifically, a point in space with coordinates $\mathbf{X} = (X, Y, Z)^T$ is mapped to the point where the line joining the point to the camera center, or center of projection, meets the image

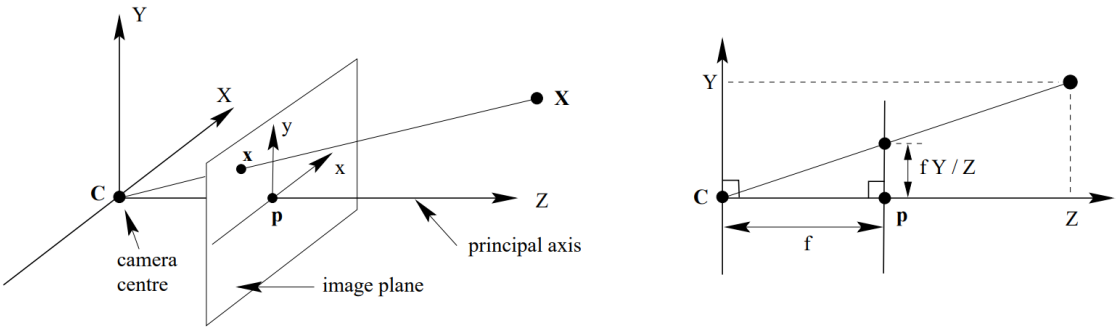


Figure 2.1: Pinhole camera model, taken from [25]

plane. The line perpendicular to the image plane passing for the camera center takes the name of principal axis, and its point of intersection with the image plane is the principal point. Using homogeneous coordinates, the pinhole camera model mapping is formalized in Equation (2.1).

$$\begin{pmatrix} fx \\ fy \\ z \end{pmatrix} = \begin{bmatrix} f_x & 0 & x_0 & 0 \\ 0 & f_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \mathbf{KX} \tag{2.1}$$

In Equation (2.1) f_x and f_x express the camera's focal length, i.e., the distance between the center of projection and the image plane. The variables x_0 and y_0 are the principal point offset, which are the principal point coordinates in the image reference frame, as presented in Figure 2.2. The focal length, principal point offset and camera's field of view are related by Equation (2.2).

$$f_x = \frac{x_0}{\tan(FOV_x/2)} \tag{2.2}$$

Equation (2.1) can be rewritten in compact form as:

$$\mathbf{x} = \mathbf{K}\Upsilon_0\mathbf{X} \tag{2.3}$$

The matrix $\mathbf{K}$, explicitly defined in Equation (2.1), is the intrinsic camera matrix. The parameters of this matrix express the geometrical parameters of the camera. The term Υ_0 expresses a rotation matrix that rotates the reference frame to have the y-axis pointing downward, as this is the convention commonly used for the image reference plane, as reported in Figure 2.2.

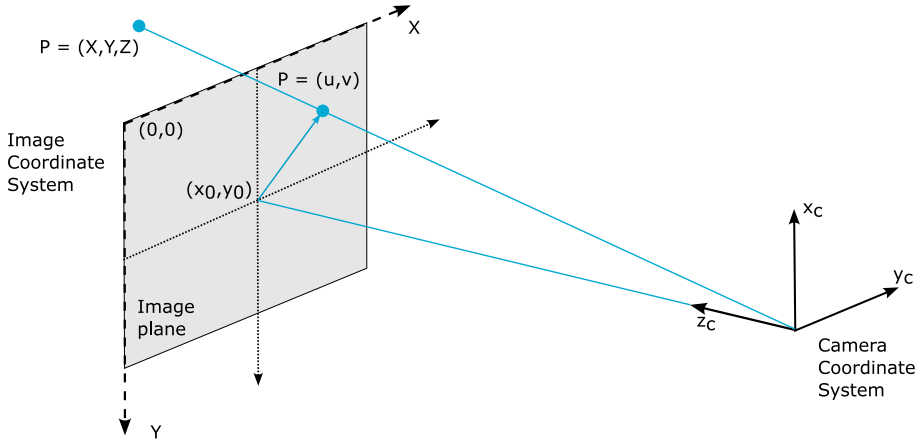


Figure 2.2: Camera coordinate system and image coordinate system

The equations presented so far stand under the assumption that the position of the real-world point is known and expressed in the camera coordinate system. Commonly, the points to be projected by the camera model are expressed in a different coordinate system, generically called world reference frame. An example of this situation is reported in Figure 2.3. It is helpful to relate the mapping going from the position in the world reference frame to the image plane in a compact expression.

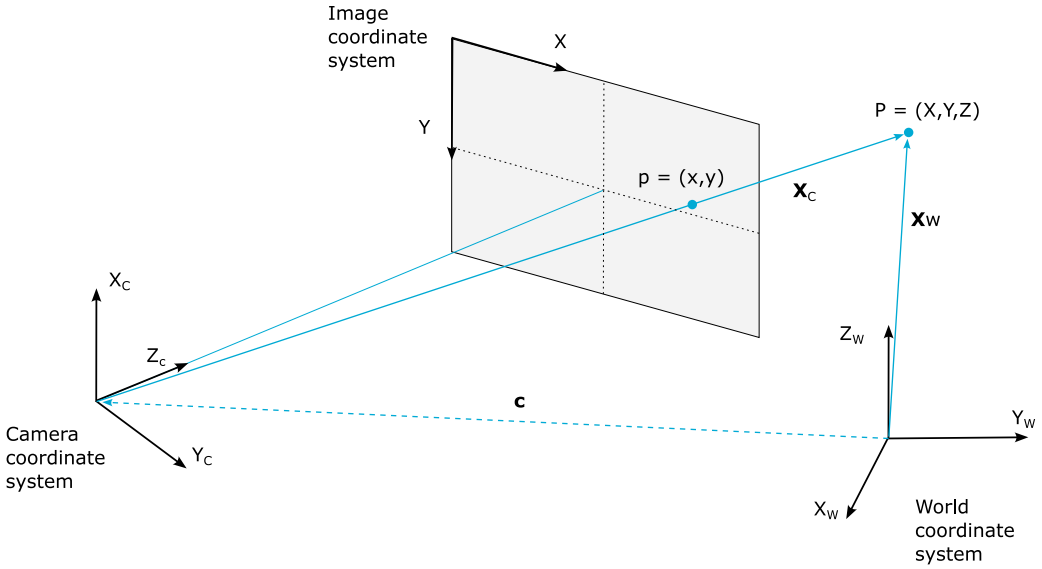


Figure 2.3: Camera coordinate system and world coordinate system

By defining a three-element column vector **c** representing the camera center position in the world reference frame and a rotation matrix **R** that transforms the world coordinates into camera coordinates, the transformation between world coordinates to camera coordinates

can be formalized as:

$$\mathbf{X}^{cam} = \mathbf{R}\mathbf{X}^{world} - \mathbf{R}\mathbf{c} \tag{2.4}$$

By defining the center of the world reference frame in camera coordinates as $\mathbf{t} = -\mathbf{R}\mathbf{c}$, and using homogeneous coordinates, Equation (2.4) can be rearranged as:

$$\mathbf{X}^{cam} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \mathbf{X}^{world} \tag{2.5}$$

Including Equation (2.5) into Equation (2.1) a mapping between a 3D point expressed in world coordinates and its position on the image plane is obtained:

$$\mathbf{x}^{image} = \mathbf{K}\Upsilon_0[\mathbf{R}|\mathbf{t}]\mathbf{X}^{world} \tag{2.6}$$

It is therefore possible to define the projection matrix $\mathbf{P}$:

$$\mathbf{P} = \mathbf{K}\Upsilon_0[\mathbf{R}|\mathbf{t}] \tag{2.7}$$

The projection matrix is the combination of the intrinsic parameters of the camera $\mathbf{K}$, representing its geometrical proprieties, and the extrinsic parameters $[\mathbf{R}|\mathbf{t}]$ that represent the camera position.

2.1.2 2D Projective transformations

Figure 2.4 shows a scenario where a set of points with world coordinates $\mathbf{X}$ is observed from two cameras with projection centers C_1 and C_2, which have different poses. That condition can occur either in the case of two different cameras at different positions or where the same camera takes two consecutive images in time with relative camera-object motion. Those two camera histories mathematical description is anallogous [25].

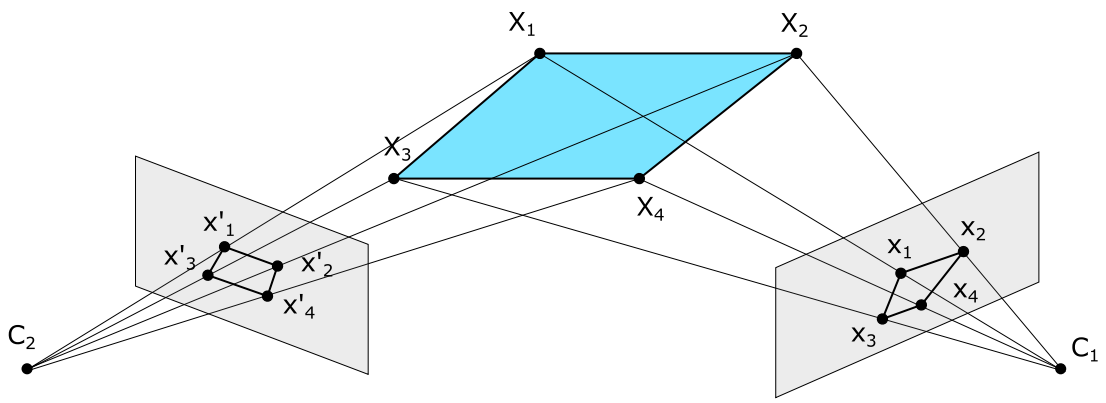


Figure 2.4: Same object multiple view

$\mathbf{x}$ and $\mathbf{x}'$ represent the points homogenous coordinates on the image plane, respectively. The linear relationship between $\mathbf{x}$ and $\mathbf{x}'$ is named projective transformation, reported in Equation (2.8).

$$\mathbf{x}' = \mathbf{H}\mathbf{x} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \mathbf{x} \tag{2.8}$$

The matrix $\mathbf{H}$ is referred to as homography matrix. The transformation is defined by eight degrees of freedom, as the matrix is defined up to a scale [18]. This reflects the fact that the transformation relates homogeneous coordindates, which are also defined up to a scale.
Projective transformations can be seen as the generalization of simpler transformations, such as similarities and affine transformations. The detailed description of those transformations is beyond the thesis scope, the reader is referred to [25] for an in depth presentation of the subject.

2.1.3 Direct Linear Transformation algorithm

Given a set of 2D to 2D point correspondences, estimating the homography between the two sets of points might be useful, as it encodes information about the relative pose of the two views [25]. Having the homography matrix 8 degrees of freedom, just four two-dimensional points are needed to estimate the transformation. A simple and effective algorithm to perform this is the Direct Linear Transformation (DLT) algorithm [25].
If two points expressed in homogeneous coordinates $\mathbf{x} = [x\ y\ z]^T$ and $\mathbf{x}' = [x'\ y'\ w']^T$ satisfy the relation $\mathbf{x}' = \mathbf{H}\mathbf{x}$ (Equation (2.8)), than the following also must hold:

$$\mathbf{x}' \times \mathbf{H}\mathbf{x} = \mathbf{0} \tag{2.9}$$

The j-th row of $\mathbf{H}$ can be denoted as $\mathbf{h}_j^T$, so that Equation (2.9) becomes:

$$\begin{pmatrix} y'\mathbf{h}_3^T\mathbf{x} - w'\mathbf{h}_2^T\mathbf{x} \\ w'\mathbf{h}_1^T\mathbf{x} - x'\mathbf{h}_3^T\mathbf{x} \\ x'\mathbf{h}_2^T\mathbf{x} - y'\mathbf{h}_1^T\mathbf{x} \end{pmatrix} = \mathbf{0} \tag{2.10}$$

Equation Equation (2.10) can be expressed as function of $[\mathbf{h}_1^T\ \mathbf{h}_2^T\ \mathbf{h}_3^T]^T$ so to define a linear

system with the components of the homography matrix as unknowns:

$$\begin{bmatrix} \mathbf{0}^T & -w'\mathbf{x}^T & y'\mathbf{x}^T \\ w'\mathbf{x}^T & \mathbf{0}^T & -x'\mathbf{x}^T \\ -y'\mathbf{x}^T & x'\mathbf{x}^T & \mathbf{0}^T \end{bmatrix} \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{pmatrix} = \mathbf{0} \tag{2.11}$$

Since two out of three are linearly independent equations, the system can be reduced to:

$$\begin{bmatrix} \mathbf{0}^T & -w'\mathbf{x}^T & y'\mathbf{x}^T \\ w'\mathbf{x}^T & \mathbf{0}^T & -x'\mathbf{x}^T \end{bmatrix} \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{pmatrix} = \mathbf{0} \tag{2.12}$$

If the system in Equation (2.12) is computed for four points it will result in a system with nine unknowns, namely the components of the homography matrix, and eight equations, thus resulting under-constrained. That highlights that only the ratio of the components of $\mathbf{H}$ is relevant. A condition to force the matrix norm to be unitary is introduced to solve the under-constraining issue.
If more than 4 point correspondences are considered, the system becomes over-determined; therefore, a least-square approach must be applied to find the solution. A practical way to solve the over-determined system is to apply the Singular Value Decomposition (SVD) to the matrix of coefficients and take the last vector of the matrix $\mathbf{V}$ as the solution.
A significant issue is overseen whenever the DLT defines the Equation (2.12) solution: the quality of the result is conditioned by the coordinate frame in which the points are expressed. This implies that some reference frames provide better results than others. As suggested by [25], a normalization to the points sets to increase the DLT algorithm's performances can be applied. The normalization consists of defining two transformations $\mathbf{T}$ and $\mathbf{T}'$ respectively for the sets of points $\mathbf{x}$ and $\mathbf{x}'$ that translate the centroid of the sets at the origin and scales them such than their average distance from the origin equals $\sqrt{2}$. Once this normalization is applied, Equation (2.12) cab be solved and then the matrix $\mathbf{H}$ denormalized. The complete DLT algorithm is summarised in Algorithm 1.

Algorithm 1 Normalized DLT for 2D homographies [25]

1: Given at least 4 2D to 2D point correspondences $\mathbf{x} \rightarrow \mathbf{x}'$
2: Compute the normalization of $\mathbf{x}$ such that $\tilde{\mathbf{x}} = \mathbf{T}\mathbf{x}$
3: Compute the normalization of $\mathbf{x}'$ such that $\tilde{\mathbf{x}}' = \mathbf{T}'\mathbf{x}'$
4: Solve the linear system in Equation (2.12) using the normalized coordinates to obtain the homography $\tilde{\mathbf{H}}$
5: Denromalize the homography matrix such that $\mathbf{H} = \mathbf{T}'\tilde{\mathbf{H}}\mathbf{T}$

2.1.4 Random Sample Consensus method

The DLT algorithm, presented in Section 2.1.3, assumes that the position of points on the image plane is not affected by noise and that outliers (elements following a different, and possibly unmodelled, error distribution [25]) don't exist in the set considered. A widespread algorithm for model fitting data containing spurious points is the RANSAC method, introduced by Fischler and Bolles in [20].
The RANSAC algorithm, summarised in Algorithm 2, starts by randomly selecting the minimum number of correspondences required for the model fitting (4 for the homography), and computes the model based on the values selected. It then uses the newly fitted model to compute the number of inliers, evaluating the distance of the points from the fitted model. This process is repeated iteratively until the exit condition is reached. The algorithm's output is the model that maximizes the number of inliers.

Algorithm 2 RANdom SAmple Consensus (RANSAC) [20]

Let A be a set of N correspondaces
while $n_{iter} < n_{max}$ and $\epsilon > \epsilon_{max}$ **do**
    Select a sample s from the set A
    Fit the model to s
    Compute the distance of all points in A from this model
    Compute the inliers set and the percentage of outliers ϵ_i
    if $\epsilon_i < \epsilon$ **then**
        Update ϵ and save model
    end if
end while

RANSAC is an iterative and non-deterministic algorithm. This implies that the result is not repeatable. It depends on the randomness of the selected points to fit the model, and its accuracy depends on the maximum number of iterations. Therefore, the number of iterations is selected high enough to ensure with a given probability p that at least one of the selected random sets does not include any outliers. The maximum number of iterations n_{max} can be computed as:

$$n_{max} = \frac{log(1-p)}{log(1-(1-\epsilon)^s)} \tag{2.13}$$

Figure 2.5 reports an example of RANSAC applied to the case of a line-fitting problem. It can be observed that the least squares fit is highly influenced by the presence of the outliers, while the RANSAC fit is more consistent with the inliers. However, it is possible to obtain an even more reliable solution to the problem by adding a step to the RANSAC algorithm. Once the RANSAC model and the inliers set are computed, applying the

least squares fit to all the correspondences in the inliers set provides the best estimate of the model, as seen in Figure 2.5. In the case of homography estimation, this procedure takes the name of Gold Standard Algorithm [25], as it provides the best estimate of the homography matrix.

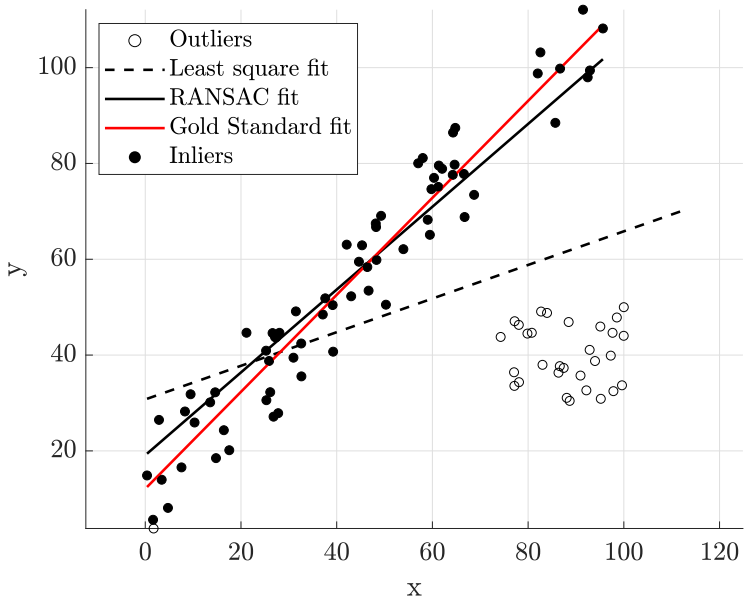


Figure 2.5: Example of line fitting with different algorithms

2.1.5 Point features

In Computer Vision, a feature is generally defined as a distinguishable element of an image. Features can be of many types, such as points, edges, circles, or colors [49]. The following sections further detail the concept of point features, also known as keypoints or corner features, introducing the most common methods to detect describe and track these features.

Features detection & extraction

Feature detection is a process that aims to identify for each image point if it corresponds to a feature. The output is a set of points coordinates, subset of the image pixels position, corresponding to each feature identified. Methods for point features detection can be divided into two categories:

- *corner detectors*, which identifies the intersection between two edges;
- *blob detectors*, which detect features by considering a supporting surrounding region.

Once the features location is identified, a neighboring region can be defined and encoded into a numerical descriptor. That process refers to as *feature extraction* or *feature description*, and it enables matching the same feature in different images without searching the whole image.
Different point features detectors and descriptor can be found in literature. A detailed overview of the most common and an evaluation of their applicability for multispectral navigation is reported in [49], highlighting a better performance of blob detectors for thermal images. Although modeled equally from a geometrical perspective, visible and thermal images present an overall difference in appearance, requiring different considerations in terms of features detection.
The selection process of the methods used in the thesis is based on a trade-off between edge detectors methods and blob detectors reported in Table 2.1. That set of algorithms includes both methods for detection only and both detection and extraction because, as explained in in the Feature tracking paragraph of Section 2.1.5, the feature descriptors are not adopted in this work.

Method	Type	Function	Reference
FAST	Corner detector	Detection	[50]
ORB	Corner detector	Detection & description	[51]
BRISK	Corner detector	Detection & description	[32]
SIFT	Blob detector	Detection & description	[35]
SURF	Blob detector	Detection & description	[5]

Table 2.1: Features detection methods evaluated in the thesis work

The detailed description of these methods is outside the scope of this work, which focuses on their application. For a detailed description of the algorithms the reader is referred to their formulation in the related papers [5, 32, 35, 50, 51].
As the feature detection process is highly dependent on the pixel intensity in the image [49], whenever a scarce contrast within the image holds the number of detected features is limited. That typically happens while processing thermal images whenever the temperature profile is almost uniform within the target and visible images whenever the target is in shadow. The image pre-processing beforehand the feature extraction through histogram equalization mitigates those issues. This process consists into splitting the image into different tiles, and then equalizing the pixel intensity so to increase the contrast and enhance the edges. In particular, the Contrast Limited Adaptive Histogram Equalization

(CLAHE) [68] algorithm successfully avoids overshooting in the contrast in sections with a uniform texture, like the background of the image. An example on the effect of CLAHE on a thermal image can be appreciated comparing Figure 2.6 and Figure 2.7.

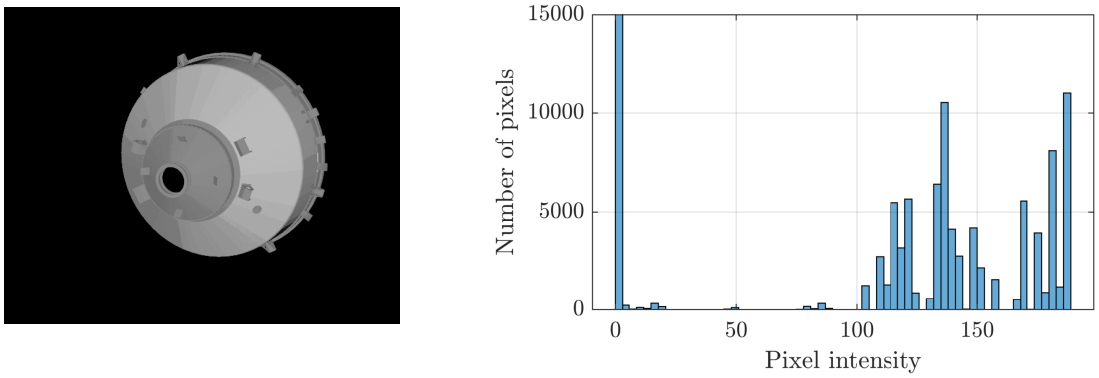


Figure 2.6: Thermal image and intensity histogram before applying CLAHE

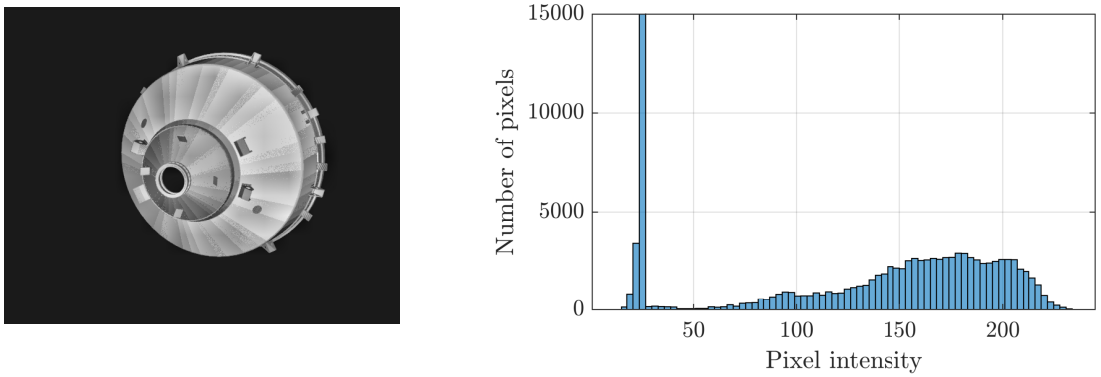


Figure 2.7: Thermal image and intensity histogram after applying CLAHE

Features tracking

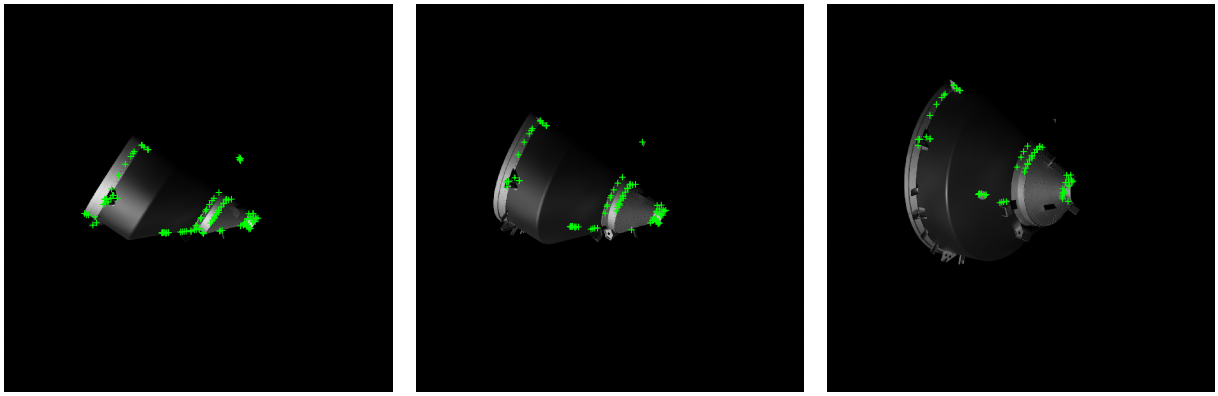
Visual navigation benefits from tracking same features in subsequent images. This problem is commonly referred to as *feature tracking*. A straightforward solution would be for each image to extract a new set of features, then match them with the original set using the descriptor information. However, this solution might result in a heavy computational burden, making it less suitable for onboard execution. In optical navigation applications, if the frame rate of acquisition of the images is in the order of seconds, the pose difference between two consecutive images is limited. This enables to use *feature tracking* techniques that extract keypoints only in the first image and track them in consecutive images searching for them in a bounded spot around the previous position. The output of each step is the field of displacement of the features between two consecutive images, known as *optical flow*.

Whenever feature tracking algorithms are adopted, two aspects have to be considered. Firstly, some features might be erroneously tracked, asking for an outlier rejection routine such as RANSAC at each step to remove spurious results. Secondly, the number of features will strictly decrease over time: that happens as the features are detected in the first image only, and over time they might either move outside the camera's field of view or get shadowed, as can be seen in Figure 2.8. To tackle those issues, the detection step shall be revisited as soon as the number of tracked key points gets below a defined threshold.

A widely used algorithm to perform feature tracking is the Lucas-Kanade (LK) tracker [36]. This algorithm associates a movement vector to every key point from its position in the first image to the position in the second image. To register the position of the key points in the two images, three assumptions hold:

- the color of a pixel, or its intensity in gray scale images, does not change over two consecutive images.
- The motion of the keypoints is limited to a bounded neighboring region around its original position.
- The keypoints are assumed to lie on rigid objects, thus moving coherently.

The research adopts the pyramidal version of the LK algorithm implemented in the MATLAB computer vision toolbox. As a detailed description of the algorithm is outside the scope of this work, which focuses on its application, the reader is referred to [4] and [9] for an overview of the different LK formulations and the implementation of the pyramidal LK respectively.



(a) Detected keypoints (n. = 115)

(b) Tracked keypoints (n. = 85) after 30 s

(c) Tracked keypoints (n. = 67) after 60 s

Figure 2.8: Features tracked with the LK algorithm in different frames taken at a frequency of 1 Hz with a relative angular rate of 1.5 deg/s

2.2 Non-linear estimation

In vision-based navigation, the features tracked on the image represent the primary source of information from the sensors. However, feature detection and tracking are always subject to a certain degree of noise, propagating into the pose information. Moreover, the information given by the position of the features at a given time is not directly influenced by the relative velocity or the angular rates, thus making their direct determination impossible without considering a sequence of measurements. This is commonly referred to as the observability problem of dynamical systems.
For the thesis work a filtering approach based on the Klaman Filter (KF) was selected because its capability to deal with nonlinear dynamics and measurements functions while maintaining a lightweight implementation. The Kalman Filter is a first-order Bayesian filter that iteratively predicts the states' estimates propagating them with a dynamical model (prediction step) to update the prediction with the information from the measurements (correction step). In the interest of brevity, the demonstration and formulation of the Kalman filter are not reported. For a detailed explanation of the method, the reader is referred to [66].
The classic formulation of the Kalman Filter works for linear systems only, while the relative navigation problem is intrinsically non-linear [44]. The filter's formulation has been extended to deal with non-linearity with the Extended Kalman Filter (EKF) and the Unscented Kalman Filter (UKF). The latter is less computationally efficient but is more accurate than the EKF, as it does not require any linearization of the dynamics. Because the thesis applies to close-proximity operations with short measurement update intervals, making the nonlinearity less problematic, the EKF was considered the most effective solution.

2.2.1 The Extended Kalman Filter

To introduce the EKF, it is first necessary to describe the real world set of non-linear differential equations describing the system:

$$\dot{\mathbf{x}} = f(\mathbf{x}) + \boldsymbol{w} \tag{2.14}$$

$$\mathbf{y} = h(\mathbf{x}) + \boldsymbol{v} \tag{2.15}$$

where $\mathbf{x}$ is the vector of system states, $\mathbf{y}$ is the vector of measurements, $f(\mathbf{x})$ is the nonlinear function of the states and $h(\mathbf{x})$ is the measurement function that maps the states into the measurements. $\boldsymbol{w}$ and $\boldsymbol{v}$ are zero mean gaussian noises with known covariance

such that:

$$\mathbf{Q} = E(\boldsymbol{w}\boldsymbol{w}^T) \tag{2.16}$$

$$\mathbf{R} = E(\boldsymbol{v}\boldsymbol{v}^T) \tag{2.17}$$

where $E(\cdot)$ denotes the expected value. $\mathbf{Q}$ and $\mathbf{R}$ are referred to as process noise matrix and measurement noise matrix respectively.
To make the classical Kalman Filter formulation applicable to non-linear problems, in the EKF the both the state and the measurement functions are linearized in the current state estimate, computing the Jacobian matrices of the two functions:

$$\mathbf{F} = \left.\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}} \tag{2.18}$$

$$\mathbf{H} = \left.\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}} \tag{2.19}$$

Further, to propagate the covariance matrix between two consecutive time steps, it is necessary to compute the state transition matrix $\mathbf{\Phi}$. This can be performed with a variety of approaches, such as finite differences, variational approach or exponential matrix. With this linearization it is possible to consider the system to be linear at each step of the filter. The steps of the Extended Kalman Filter are reported in Algorithm 3.

Algorithm 3 Extended Kalman Filter

1: $\mathbf{x}_k^- = \mathbf{\Phi}(t_k, t_{k-1})\mathbf{x}_{k-1}^+$ ▷ State & covariance propagation
2: $\mathbf{P}_k^- = \mathbf{\Phi}(t_k, t_{k-1})\mathbf{P}_{k-1}^+\mathbf{\Phi}(t_k, t_{k-1}) + \mathbf{Q}$
3: $\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^-$ ▷ Innovation & covariance residual computation
4: $\mathbf{S}_k = \mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}$
5: $\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^T\mathbf{S}_k^{-1}$ ▷ Kalman gain
6: $\mathbf{x}_k^+ = \mathbf{x}_k^- + \mathbf{K}_k\mathbf{d}_k$ ▷ State & covariance update
7: $\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^-$

2.2.2 The Multiplicative Extended Kalman Filter

Since the thesis aims at investigating the 6 degrees of freedom navigation performances the filter shall be able to estimate both the center of mass position and the relative attitude between the two spacecrafts.
The quaternion is a widely used parametrization for attitude estimation, as it is the singularity-free representation with the lowest dimensionality [37]. However, the normalization constraint might cause issues in the standard formulation of the EKF with the additive representation of the error (Algorithm 3, line 6). The Multiplicative EKF, pro-

posed in [30], successfully solves this issue. The underlying principle of the MEKF is to replace the full quaternion with a three-element vector representing the local representation of attitude errors while keeping track of a unit quaternion representing the global attitude. As a consequence, the additive state correction is substituted by a quaternion multiplication:

$$\mathbf{q}^{+} = \delta\mathbf{q}(\mathbf{a}_p) \otimes \mathbf{q}^{-} \tag{2.20}$$

where $\delta\mathbf{q}$ is the unit quaternion error, $\mathbf{a}_p$ is the three element attitude error state, and $\mathbf{q} = [q_1\ q_2\ q_3\ q_4]^T$, is the reference quaternion with q_4 as scalar part. This method solves the issues related to an additive representation of the error and reduces the number of states of the system to its degrees of freedom, providing a more compact formulation.
In this work, the parametrization used for the three-element state vector is the Modified Rodriguez Parameters (MRP) proposed in [59]. The direct and inverse mapping from quaternion to MRP is reported in Equations (2.21) and (2.22).

$$\mathbf{a}_p = \frac{4}{1+q_4}[q_1\ q_2\ q_3]^T = \frac{4}{1+q_4}\bar{\mathbf{q}} \tag{2.21}$$

$$\mathbf{q} = \frac{1}{16 + \mathbf{a}_p^T\mathbf{a}_p}\begin{bmatrix} 8\mathbf{a}_p \\ 16 - \mathbf{a}_p^T\mathbf{a}_p \end{bmatrix} \tag{2.22}$$

where $\bar{\mathbf{q}}$ indicates the vectorial part of the quaternion. From Equation (2.21) it can be deduced that the MRP are not singularity free. However, this problem does not arise in the MEKF, as the three-element error state is set to zero at each reset step of the filter. This is consequential to the assumption that the filter provides the best estimate. After the update step, the estimated quaternion is assumed to be equal to the true attitude, having a quaternion error equal to $\delta\mathbf{q} = [0\ 0\ 0\ 1]^T$ and consequentially $\mathbf{a}_p = [0\ 0\ 0]^T$. Having the scalar term of the quaternion unitary allows one to avoid the singularity condition of the MRP.

3

Visual Navigation Filter

In the present chapter, the visual navigation pipeline proposed in the thesis is presented, utilizing the tools and techniques detailed in Chapter 2 as the foundation for the work. The algorithm aims at tracking the relative attitude and position (pose) of the target with respect to the chaser using a visible and a thermal monocular camera measurements only. The core of navigation pipeline is an Extended Kalman Filter that provides the pose estimate fusing the information coming from the two sensors. The raw images acquired by the cameras are not fed directly into the filter, but are elaborated in an Image Processing step, which applies pre-processing (if needed) and tracks the relevant point features. The original aspect of the proposed navigation algorithm stays in the filter's innovation computation. The image features and the relative landmark position represent the filter's measurements and pseudo-measurements of the filter. The landmarks position results from the projection of a wireframe model of the target on the image plane according to the available state estimate, thus representing a measure of it. The update step of the filter is then able to correct the pose estimate without reconstructing the pose from the images with algorithms such as the PnP. Additionally, an online estimation of the measurement noise covariance matrix is implemented to avoid modeling errors and adjust for the sensors performance changes due to environmental conditions.
A pseudo-code description is presented in Algorithm 4, where the implementation of the process is reported. For the sake of clarity a block diagram of the pipeline is also reported in Figure 3.1.

Algorithm 4 Navigation Filter

Input: Visible and/or thermal image from navigation cameras

1: Pre-process the thermal images (histogram equalization)
2: Track the features on the images at t_k given their position in the previous measurements t_{k-1}
3: Propagate the states and the associate covariance from t_{k-1} to t_k
4: Project the model landmarks according to the pose provided by the propagated states
5: Compute the filter's innovation: image features position minus landmarks position
6: Compute the Kalman gain and perform the state update (Algorithm 3)
7: Update the estimate of the measurement noise covariance

Throughout this chapter, the visual navigation filter is presented to provide a detailed insight into the proposed navigation solution. After an introduction on the adopted reference frames, the dynamical model employed by the filter to propagate the states is discussed. The presentation of the measurement model is splitted between the description of the measurement function defined to compute pseudo-measurements and the Image Processing pipeline exploited to detect and track features. Finally, the methods used to estimate the measurement noise covariance matrix and remove outliers are described.

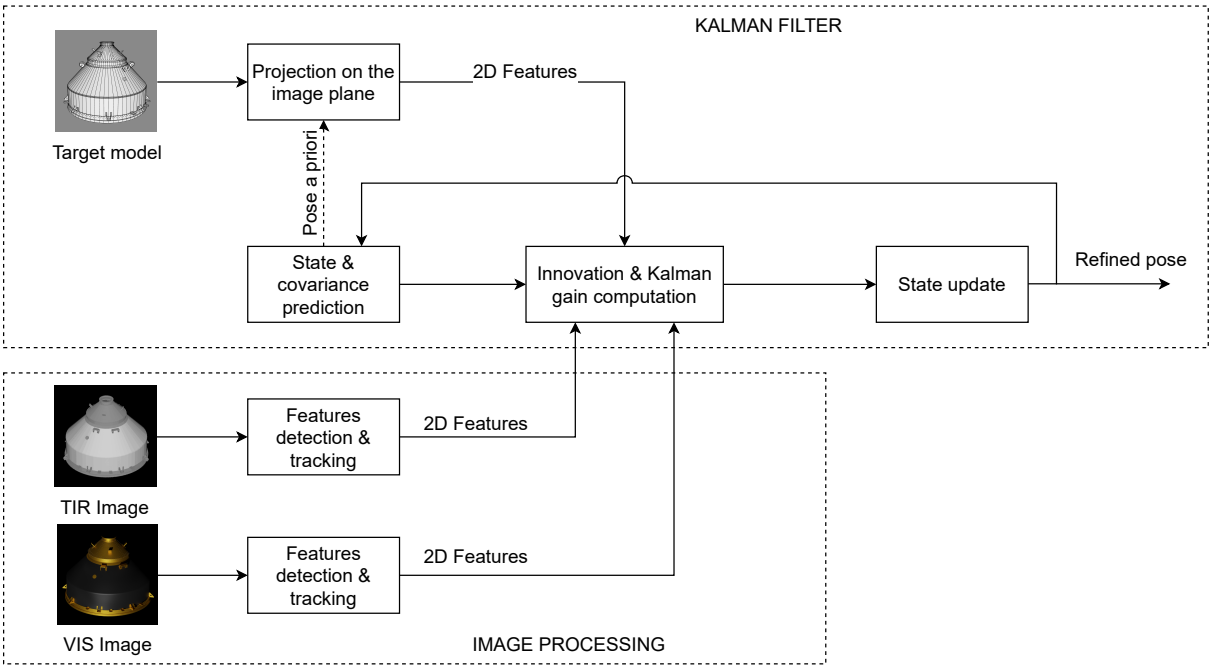
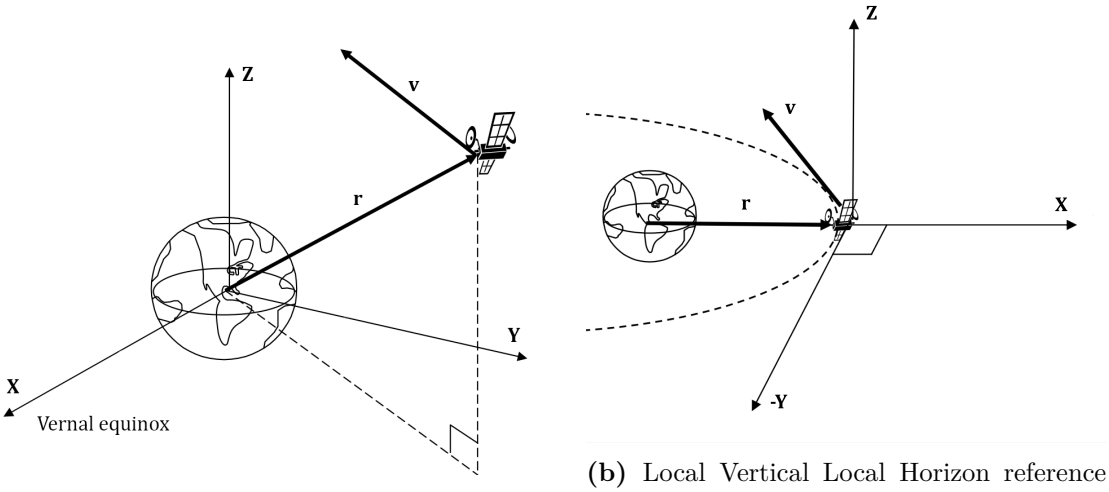


Figure 3.1: Visual navigation pipeline architecture

3.1 Reference frames

The fundamental reference frames used in the thesis are hereafter presented.
The spacecraft absolute position and attitude are expressed in Earth Centered Inertial frame $\mathcal{I}$ (Figure 3.2a).
The origin of the Local Vertical Local Horizon ($\mathcal{L}$) frame is the barycenter of the spacecraft, the x axis is directed from the spacecraft radially outward, the z axis is aligned with the normal of the orbit plane and positive along the direction of the angular momentum. The y-axis completes the right-handed triad (Figure 3.2b).

The Body frame is instead attached to the spacecraft, having the center in the center of mass of the body. The body frame ($\mathcal{B}$) axes are aligned with the principal axis of inertia of the spacecraft.
Finally the sensor coordinate system, the camera frame ($\mathcal{C}$), shall be introduced to de-



(a) Earth Centered Inertial reference frame

(b) Local Vertical Local Horizon reference frame

Figure 3.2: ECI and LVLH coordinate systems, taken from [3]

scribe the position of external elements with respect to the sensor. This coordinate system is described in detail in Section 2.1.1.
To express in which coordinate system a vector is described, its reference letter is added as subscript. In the case of the LVLH and Body reference frames, as they can be applied both to the chaser and the target, the lowercase letters c and t are added to specify the origin of the system. An example of this notation is reported in Equation (3.1).

$$\mathbf{r}_C = \mathbf{A}_{C/Bc}\mathbf{A}_{Bc/I}\mathbf{r}_I \tag{3.1}$$

The generic vector $\mathbf{r}$ is mapped from the Earth Centered Inertial to the Camera frame by mean of two consecutive rotations expressed in terms of Direct Cosine Matrices $\mathbf{A}$. For elements expressing the transformation between the two reference frames, such as rotation matrices or quaternions, the convention used is the one reported in Equation (3.1). $\mathbf{A}_{Bc/I}$ represents the chaser body frame (Bc) expressed in the inertial reference frame (I), and $\mathbf{A}_{C/Bc}$ represents the camera frame expressed in the chaser body reference frame.

3.2 Dynamical model

As the filter estimates the 6 degrees of freedom pose of the target, the dynamical model includes both the translational and the rotational dynamic of the relative motion. Specifically, the states of the filter are comprised of the position and velocity of the target in the chaser LVLH frame, a three-element attitude error expressed in MRP and the angular velocities of the target body frame expressed in the chaser body frame. In the present

section the state propagation model is detailed. During the discussion, to express the target position in the chaser's body frame ($\mathbf{x}_{Lc} = [x_{Lc}\ y_{Lc}\ z_{Lc}]$) and the relative attitude ($q_{Bt/Bc}$), the subscripts indicating the reference frames are avoided for the sake of brevity.

3.2.1 Translational dynamic

The relative motion between two spacecrafts is an intrinsically non-linear problem. An extensive derivation of the differential equation describing the target motion expressed in the chaser LVLH frame is reported in [3], resulting in the general nonlinear equations of relative motion:

$$\begin{aligned}
\ddot{x} - 2\dot{f}_c\dot{y} - \ddot{f}_c y - \dot{f}_c^2 x &= -\frac{\mu(r_c + x)}{[(r_c + x)^2 + y^2 + z^2]^{3/2}} + \frac{\mu}{r_c^2} + d_R \\
\ddot{y} + 2\dot{f}_c\dot{x} + \ddot{f}_c x - \dot{f}_c^2 y &= -\frac{\mu y}{[(r_c + x)^2 + y^2 + z^2]^{3/2}} + d_T \\
\ddot{z} &= -\frac{\mu z}{[(r_c + x)^2 + y^2 + z^2]^{3/2}} + d_N
\end{aligned} \tag{3.2}$$

where f_c and r_c represent the true anomaly and the radius of the chaser respectively, and $[d_R, d_T, d_N]$ the differential perturbing acceleration in the radial, tangential and normal directions.

In the context of spacecraft rendezvous, Clohessy and Wiltshire [15] proposed a linear form of the equations of relative motion by neglecting all perturbations considering only the first order terms of the Taylor expansion of Equation (3.2) [58]. This model of motion is here selected for the translational dynamic of the filter. The selection of such a simplified model is justified by the short measurements update intervals of close proximity operations (in the order of seconds) and by the quasi-circular orbit of space debris in LEO orbits. In case those assumption are not applicable more complex models such as the Yamanaka-Ankersen [62] are advised. The Clohessy-Wiltshire (CW) linearized equations of unperturbed relative motion are reported in Equation (3.3).

$$\begin{aligned}
\ddot{x} - 2n\dot{y} - 3n^2 x &= 0 \\
\ddot{y} + 2n\dot{x} &= 0 \\
\ddot{z} + n^2 z &= 0
\end{aligned} \tag{3.3}$$

where n indicates the mean motion of the chaser. The state transition matrix of the CW equations can be computed analitically and is reported in Equation (3.4), where c_{nt} and s_{nt} indicates $cos(nt)$ and $sin(nt)$ respectively [3].

$$\Phi_{CW} = \begin{bmatrix} 4-3c_{nt} & 0 & 0 & \frac{s_{nt}}{n} & \frac{2}{n}-\frac{2c_{nt}}{n} & 0 \\ -6nt+6s_{nt} & 1 & 0 & -\frac{2}{n}+\frac{2c_{nt}}{n} & \frac{4s_{nt}}{n}-3t & 0 \\ 0 & 0 & c_{nt} & 0 & 0 & \frac{s_{nt}}{n} \\ 3ns_{nt} & 0 & 0 & c_{nt} & 2s_{nt} & 0 \\ -6n+6nc_{nt} & 0 & 0 & -2s_{nt} & -3+4c_{nt} & 0 \\ 0 & 0 & -ns_{nt} & 0 & 0 & c_{nt} \end{bmatrix} \tag{3.4}$$

3.2.2 Rotational dynamic

The target absolute angular accelerations cannot be determined in an uncooperative scenario; the model here used for attitude dynamics assumes constant small relative angular velocities, as proposed in [24, 53]. A quaternion parametrization is adopted, formalized in Equations (3.5) and (3.6).

$$\dot{\mathbf{q}} = \frac{1}{2}\Omega(\boldsymbol{\omega})\mathbf{q} = \begin{bmatrix} \boldsymbol{\omega} \\ 0 \end{bmatrix} \otimes \mathbf{q} \tag{3.5}$$

$$\dot{\boldsymbol{\omega}} = \mathbf{0} \tag{3.6}$$

where $\mathbf{q} = [q_1\ q_2\ q_3\ q_4]^T$, with q_4 as scalar part, is the target body frame expressed in the chaser body frame, and $\boldsymbol{\omega}$ the relative angular velocities vector. $\boldsymbol{\Omega(\omega)}$ is a 3×4 matrix defined as:

$$\Omega(\boldsymbol{\omega}) = \begin{bmatrix} -[\boldsymbol{\omega}\times] & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0 \end{bmatrix} \tag{3.7}$$

where $[\,\cdot\,\times]$ indicates the skew-symmetric matrix. As the filter exploits the multiplicative representation of the quaternion error (Section 2.2.2), it is necessary to derive the attitude dynamical model as function of the three-element error vector $\mathbf{a}_p$ expressed in MRP. This

can be obtained by substituting Equation (3.5) in the time derivative of Equation (2.21):

$$\begin{aligned}\dot{\mathbf{a}}_p &= \frac{4}{1+q_4}\dot{\bar{\mathbf{q}}} - \frac{4}{(1+q_4)^2}\dot{q_4}\bar{\mathbf{q}} \\ &= \Big(-\frac{1}{2}[\boldsymbol{\omega}\times] + \frac{1}{8}\boldsymbol{\omega}\cdot\mathbf{a}_p\Big)\mathbf{a}_p + \Big(1 - \frac{1}{16}\mathbf{a}_p^T\mathbf{a}_p\Big)\boldsymbol{\omega} \\ &= -\frac{1}{2}[\boldsymbol{\omega}\times]\mathbf{a}_p + \boldsymbol{\omega}\end{aligned} \tag{3.8}$$

Equations (3.6) and (3.8) represent the differential equations of the attitude dynamical model. In this case, the State Transition Matrix $\boldsymbol{\Phi}_{att}$ cannot be computed analytically but requires numerical computation. The approach suggested in [59] is to use the exponential matrix, and, as no difference in terms of accuracy was found comparing this method with the variational approach, it is used for the computation of the STM.

The complete rotational-translationa propagation model can be expressd in compact form as:

$$\begin{pmatrix}\dot{\mathbf{x}}\\ \ddot{\mathbf{x}}\\ \dot{\mathbf{a}}_p\\ \dot{\boldsymbol{\omega}}\end{pmatrix} = \begin{bmatrix}\boldsymbol{\Phi}_{CW} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Phi}_{att}\end{bmatrix}\begin{pmatrix}\mathbf{x}\\ \dot{\mathbf{x}}\\ \mathbf{a}_p\\ \boldsymbol{\omega}\end{pmatrix} \tag{3.9}$$

3.3 Measurement model

The measurement model is a function that transforms the filter states to the estimates of the measurements coming from the sensors. In the here treated context, this implies mapping the landmarks of the target known 3D model, expressed in the target body frame, to their estimated position on the image plane. The measurement function building can be conceptually divided in three steps, namely the transformation of the features position from the target body frame to the camera frame (h_1), a second step of projection of the points on the image plane (h_2), and finally the conversion from homogeneous to Cartesian coordinates (h_3).
The first step is formalised in Equation (3.10).

$$\mathbf{f}_C = h_1(\mathbf{x}_{Lc}, \mathbf{a}_p) = \mathbf{A}_{C/Bc}(\mathbf{A}(\mathbf{q}(\mathbf{a}_p)\otimes\mathbf{q}_{Bt/Bc})^T\mathbf{f}_{Bt} + \mathbf{A}_{Bc/I}\mathbf{A}_{I/Lc}\mathbf{x}_{Lc}) - \mathbf{A}_{C/Bc}\mathbf{t}_{Bc} \tag{3.10}$$

where $\mathbf{f}_C$ is a generic landmark position in the camera reference frame, $\mathbf{f}_{Bt}$ its position in the target body frame and $\mathbf{t}_{Bc}$ the position of the camera in the chaser body frame. The transformation matrices $\mathbf{A}_{Bc/I}$ and $\mathbf{A}_{I/Lc}$ can be determined because the position and

attitude of the chaser in the inertial frame are assumed to be known. In real applications, the DCM $\mathbf{A}_{C/Bc}$ and the translation vector $\mathbf{t}_{Bc}$ have different values for different cameras, as they are necessarily placed in different positions onboard the chaser. However, within the current application both the thermal and the visible cameras are assumed to lie in the center of the chaser mounted with the same attitude. This assumption is not realistic but does not entail any difference to the navigation pipeline.
The MRP attitude error $\mathbf{a}_p$, used to retrieve the estimated attitude of the chaser in the attitude reference frame, and the target position in the chaser LVLH frame $\mathbf{x}_{Lc}$ are the states dependent on the measurements. As expected, the velocity and angular rate do not appear, as they do not provide any information on the features position at a given time.
The second step corresponds to the projection of the points to the image plane according to the pinhole camera model described in Section 2.1.1:

$$\mathbf{f}^h = h_2(\mathbf{f}_C) = \mathbf{K}\mathbf{f}_C \tag{3.11}$$

where $\mathbf{f}^h = [f_x^h \ f_y^h \ f_z^h]^T$ indicates the feature position on the image plane expressed in homogeneous coordinates. As the visible and thermal cameras are assumed to be coincident, the intrinsic camera matrix is the only factor that distinguishes the two sensors within the context of the measurement function.
The final step consists in the non-linear transform between the homogeneous and Cartesian coordinates:

$$\mathbf{f} = h_3(\mathbf{f}^h) = \begin{cases} \dfrac{f_x^h}{f_z^h} \\ \dfrac{f_y^h}{f_z^h} \end{cases} \tag{3.12}$$

Combining the defined equations according the definition of Equation (2.15), and defining the output pseudo-measurements $\mathbf{y}$ as the landmark position $\mathbf{f}$, the full measurement function can be defined as:

$$\mathbf{y} = h_3(h_2(h_1(\mathbf{x}_{Lc}, \mathbf{a}_p))) \tag{3.13}$$

The resulting measurement function is non-linear, therefore the Jacobian of the function has been computed with respect to all the filter states to be compatible with the EKF formulation. Since it is not function of the velocity or the angular rates, the columns of the Jacobian associated to those states will be zeros.

3.4 Image Processing pipeline

The scope of the image processing pipeline is to provide the image features location and their relation to the wireframe model points. A flow diagrams of the IP pipeline is reported in Figure 3.3.

The features position on the image acquired at time t_k are the output of the Lucas Kanade tracker, using the results obtained at time t_{k-1} as initial condition. As the features are tracked along different images, the information about their correlation with the model landmarks is maintained, avoiding the need of a matching process at each filter step. To remove either spurious associations or errors collected during the tracking step, a RANSAC routine is added to remove possible outliers. If the number of features after the outlier rejection is over a defined threshold k, then the IP step can be considered satisfactory and the information acquired is used in the filter to refine the estimated pose.

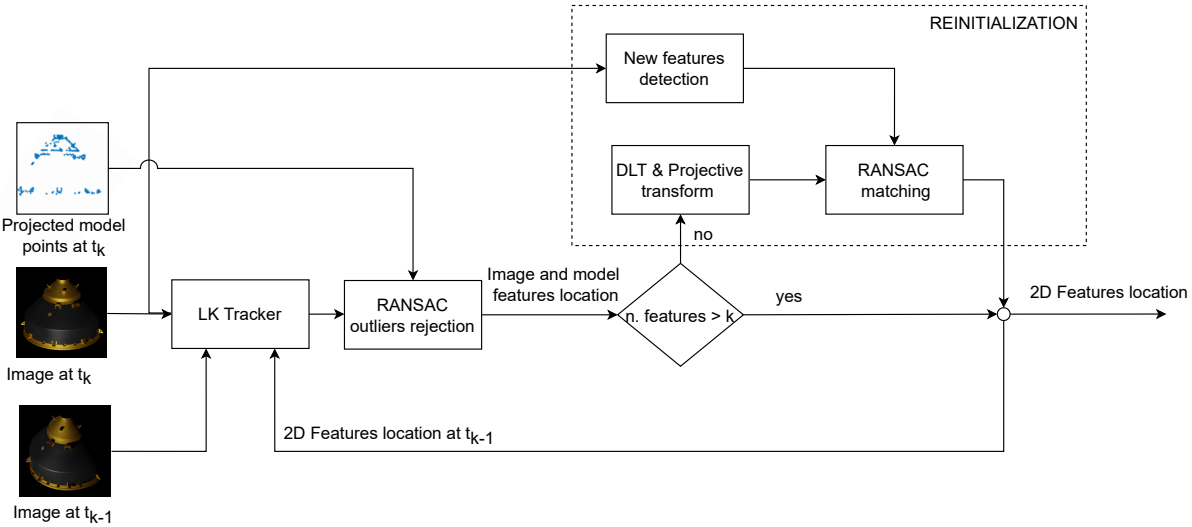


Figure 3.3: Image processing pipeline architecture

However, as explained in Section 2.1.5, the number of features strictly decreases over time, requiring a re-initialization step.

In the re-initialization the problem of redefining the correlations between image features and model landmarks arises. A possible solution is to perform a brute-force RANSAC association by randomly sampling the entire set of image features and model points until the correct correspondences are extracted and the associations are computed. Given the large number of features, the number of iterations required to ensure a good probability of success would make the algorithm prohibitive for real-time applications.

The solution here proposed exploits the information about the image-to-map correlation

available before the re-initialization process to limit the brute-force RANSAC search region. The steps of the implemented solution are described in pseudocode in Algorithm 5. Given the output of the RANSAC outlier rejection routine, the homography matrix mapping the model landmarks to the respective image features is computed. To combine those two steps is equivalent to applying the Golden Standard algorithm (Section 2.1.4), thus providing the best estimate of the projection of the landmarks onto image features. At this point, it is possible to use the retrieved projective transformation matrix to map all model landmarks to where their corresponding image feature is expected to be located. Once detected the new features from the image, it is reasonable to assume that the associated landmarks are restricted to a limited region about their location. When performing the RANSAC polling, only the landmarks within a certain distance d_{max} from the respective extracted feature are considered. The more the search region dimension is extended, the more the method is robust to estimation errors, although increasing the computational burden. The search region is therefore to be tuned to reach the desired trade-off between robustness and computational time.

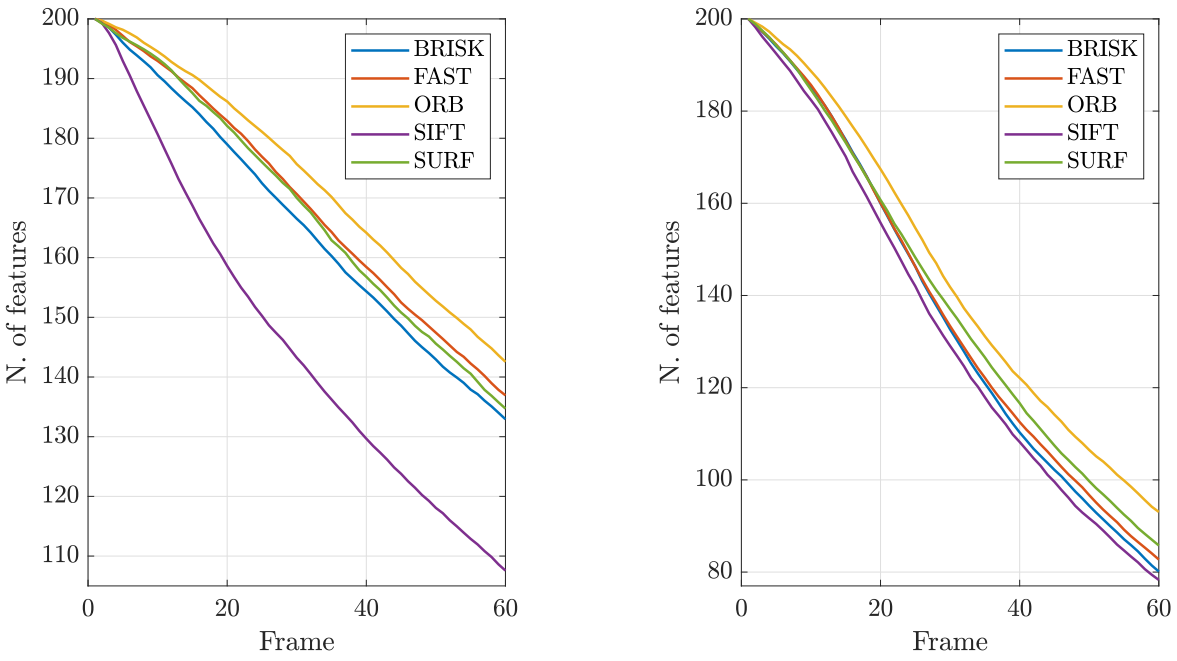
Algorithm 5 IP re-initialization algorithm

Given: a vector of image features locations $\mathbf{f}_a^i$ and the associated projected model points locations $\mathbf{f}_a^m$, subset of all model landmarks on the image plane $\mathbf{F}_a^m$
With DLT, compute the homography s.t. $\mathbf{f}_a^i \simeq \mathbf{H}\mathbf{f}_a^m$
Trasform all the model points s.t. $\mathbf{F}_b^m = \mathbf{H}\mathbf{F}_a^m$
Detect a new set of features $\mathbf{f}_b^i$ from the image
while $n_{iter} < n_{max}$ **do**
    Randomly select 4 elements from $\mathbf{f}_b^i$
    Randomly select 4 elements from the subset of $\mathbf{F}_b^m$ s.t. $\| \mathbf{F}_b^m - \mathbf{f}_b^i \| \leq d_{max}$
    Compute the homography matrix between the selected points and select the inliers $\mathbf{f}_b^m$ of the model
    Count the number inliers ε of the fitted model
    if $\varepsilon > \varepsilon_{max}$ **then**
        Update ε_{max} and save $\mathbf{f}_b^i$ and $\mathbf{f}_b^m$ as the re-intialized features
    end if
end while

The IP pipeline is almost identical for the visible and thermal spectra. The only difference is the application of the CLAHE contrast equalization to the thermal image. Because of the overall lack of contrast of thermal images, its enhancement is beneficial for the features identification process [49].

3.4.1 Feature detector trade-off

The selection of the feature detector is based on a trade-off performed between different methods. The methods investigated include both blob detectors (SIFT, SURF) and corner detectors (BRISK, FAST, ORB). For all those methods, the formulatio implemented in the Computer Vision toolbox of MATLAB is here adopted.
The trade-off is performed against merit-criteria tailored for the thesis applications. As all methods can detect a satisfactory amount of features, given the suitable tailoring of their parameters, they are evaluated on the quality of the features detected in terms of tracking: the more a method can detect features that are reliably tracked along different frames (moving coherently with the associated point on the rigid body) the less re-initialization steps are required, thus providing a positive contribution to the navigation pipeline timeliness. To achieve this evaluation, the extracted features are tracked along 60 consecutive frames, eliminating outliers at each step. This procedure is executed 100 times on a database of 400 images, starting each run with a different initial frame to avoid test the methods on different point of views of the target.
The averaged results obtained over 100 runs are reported for the different methods in Figure 3.4.



(a) Visible images

(b) Thermal images

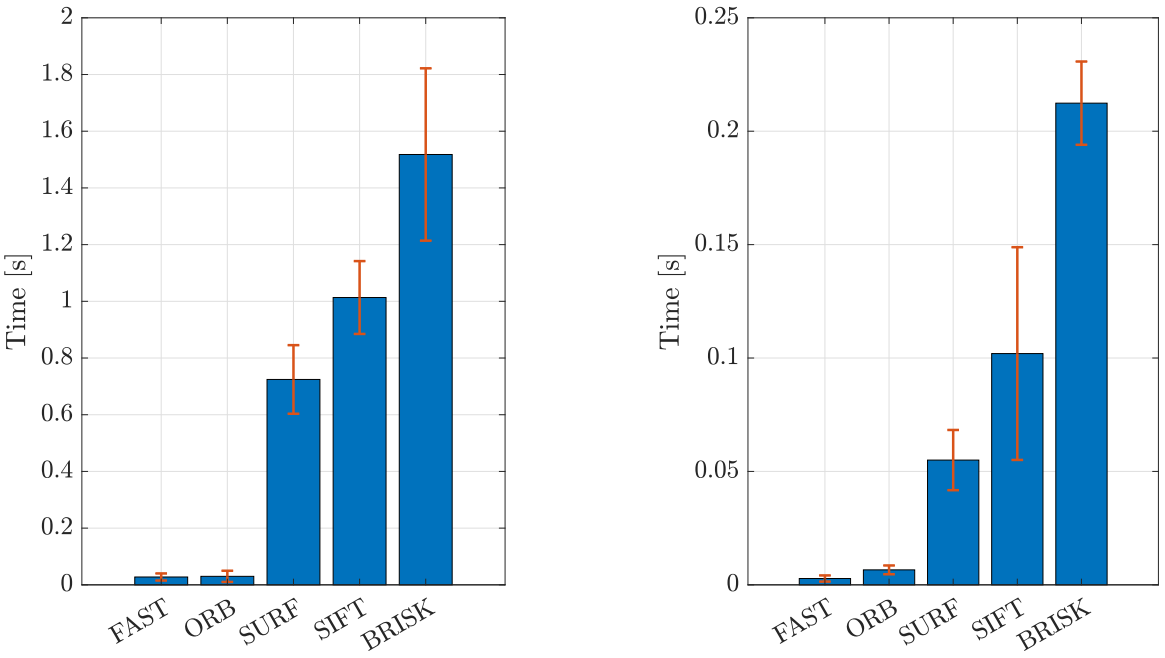
Figure 3.4: Number of features tracked along different frames for visible and thermal images

For both visible and thermal images the best method results to be ORB, as it identifies

more reliable features traceable for longer times. It can also be observed an overall difference in performances between the visible and thermal case, as TIR images present a faster feature drop rate. At a first analysis this results seemed counter-intuitive as in the thermal images the shadowing problem does not arise, thus it is expected to maintain a higher number of features over time with respect the VIS images, which are affected by that problem. The reason behind the TIR spectrum lower performances of the TIR spectrum is the spurious association of the features to the model points after the detection, given by the higher noise level in the image and the adopted simplified thermal model. Along the tracking process this outliers are then rejected, justifying the results obtained in Figure 3.4b.

The computation time required for the feature detection process is a second merit-criterion considered because a lightweight method would ease future onboard testing of the algorithm. Figure 3.5 shows results for both the visible and thermal case. In both spectra the FAST and ORB methods outperform the others, making the ORB detection the overall best solution for this work application.

It should be noted that the trade-off performed is specific to the application and does not generally apply as a benchmark for feature detector methods.



(a) Visible images

(b) Thermal images

Figure 3.5: Average time required on the feature detection process for the visible and thermal images over 100 images

3.5 Measurement noise matrix adaptation

As described in Section 3.3 the features' position on both the visible and thermal images are the measures of the filter. In the framework of the Kalman filter it is necessary to accurately model the measurement noise to avoid degradation of the filter performances or divergence from the ground truth. For navigation cameras this can result in a challenging task, as the noise associated to the detected features is influenced both by the quality of the sensor and by the external environmental conditions; aspects such as either the illumination condition or the chaser-target distance could in fact vary the noise associated to the image features.
To tackle this issue the residual-based adaptive estimation of the measurement noise covariance matrix $\mathbf{R}$ proposed in [2] is implemented.
The filter residual is defined as the difference between the measurements and the pseudo-measurements of the updated state:

$$\boldsymbol{\varepsilon}_k = \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^+ \tag{3.14}$$

As demonstrated in [2] the estimate of the residual covariance can be expressed as:

$$\hat{\boldsymbol{C}}_k = E(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^T) = \hat{\mathbf{R}}_k - \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T \tag{3.15}$$

Indicating with $\hat{\cdot}$ the estimated values. The definition for the estimate of $\mathbf{R}$ at each step becomes:

$$\hat{\mathbf{R}}_k = E(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^T) + \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T \tag{3.16}$$

The expected value of the residual covariance $E(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^T)$ can be approximated by averaging $\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^T$ over a sliding window of dimension N as:

$$E(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^T) = \frac{1}{N} \sum_{k=1}^{N} \boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^T \tag{3.17}$$

To avoid saving all residuals over the sliding window, a forgetting factor α is introduced to adaptively estimate the measurement noise covariance matrix, rewriting Equation (3.16) as:

$$\hat{\mathbf{R}}_k = \alpha \mathbf{R}_{k-1} + (1 - \alpha)(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^T + \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T) \tag{3.18}$$

The lower the value of α, the more the estimation of $\mathbf{R}$ depends on the current residual, although the estimate would be subject to fluctuations due to noise within the residuals. The forgetting factor should then be tuned to obtain the desired results from the

adaptation.

3.6 Outliers rejection

Although an outlier rejection method is already introduced in the Image Processing pipeline, it is convenient to discard the noisiest measurements to enhance the performances and the stability [12].
As it is assumed that the measurements are affected by known Gaussian noise, a null-hypothesis test is performed to check whether the measurements are in accordance with the assumed model. As highlighted in [59], the square of the Mahalanobis distance (γ_k) of the innovation can be exploited as performance metric for the null-hypothesis:

$$\gamma_k = M_k^2 = \mathbf{d}_k^T \mathbf{S}_k^{-1} \mathbf{d}_k \tag{3.19}$$

where $\mathbf{d}_k$ is the innovation, $\mathbf{S}_k$ its covariance matrix and M_k the Mahalanobis distance. The innovation is defined as:

$$\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^- \tag{3.20}$$

And the associated covariance $\mathbf{S}_k$ as:

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R} \tag{3.21}$$

Under the assumption that the null-hypothesis is true, i.e.the error is Gaussian distributed, γ_k should be Chi-square distributed with degrees of freedom equal to the dimensionality of the innovation vector. To remove possible outliers a gating method is applied, excluding all measurements which are bigger than a threshold χ_α defined so that Equation (3.22) is true.

$$P(\gamma_k > \chi_\alpha) = \alpha \tag{3.22}$$

Equation (3.22) expresses that the probability of a randomly selected γ_k to be higher than χ_α is equal to α. The value of α is selected to be 0.05. Such value ensures that the noisiest measurements are removed without excluding inliers.
That approach for outliers rejection enables both to remove possible outliers and to remove from the measurement vector the most noisy values, thus enhancing the filter's performances.

4

Testing Framework

All tests performed are based on MATLAB routine executed on a Desktop PC. The images output of the thermal sensors are the result of an a priori rendering process both for the visible and the thermal case.
To provide a detailed overview of the testing framework, the target debris VESPA [1] is presented, describing the model used to generate the landmarks. Subsequently, the rendering tool used to generate the synthetic images is detailed. Different study cases are identified from a brief discussion on the rendered synthetic images, varying the parameters that primarily influence the sensors' output. Finally, the ground truth used for relative dynamics and the filter's figures of merit are presented.

4.1 Vega Secondary Payload Adapter

VESPA is a payload adapter that enables the Vega rocket embarking and deploying different payloads in different orbits. It was first used on Vega's second flight (VV02) in 2013, deploying the Proba-V and other two payloads in two different orbits [1]. A pre-launch picture of Proba-V mounted on VESPA is shown in Figure 4.1.

Figure 4.1: Proba-V mounted on VESPA, from [1]

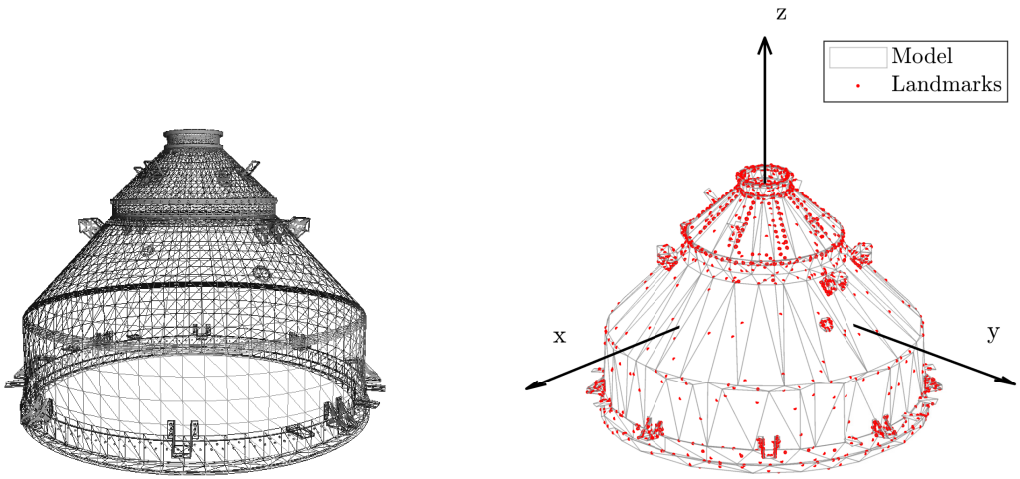
The VESPA properties are reported in Table 4.1. The VESPA debris is of interest both for the ClearSpace-1 mission [8], aimed to de-orbit the adapter, and for the e.Inspector mission, aimed at de-risking optical navigation technologies for uncooperative targets and collecting relevant information on the current debris status to support its removal operations. Because of its dimension, VESPA is considered an Intact Derelict Object, making its removal critical for space sustainability [33].

Parameter	Value	Unit
Payload interface diameter	937	mm
Launcher interface diameter	2105	mm
Height	2867	mm
VESPA mass	262	kg
Maximum main payload mass	1000	kg
Maximum secondary payload mass	200	kg

Table 4.1: VESPA properties [1]

4.1.1 VESPA wireframe model

As specified in Section 3.3, the filter uses a wireframe model of the target to compute the pseudo-measurements of the states, projecting the model features onto the image plane. A STereolithograpfy File (STL) containing a detailed model of VESPA is used to define the model landmarks. As shown in Figure 4.2b, a landmark is generated for each face of the model. If the landmarks were to be generated from a highly detailed model (Figure 4.2a), their elevated number would compromise the filter's efficiency. For this reason, the target's original model is downscaled to avoid this issue. The detailed model, the downscaled model, and the landmarks can be observed in Figure 4.2.
Figure 4.2 allows evaluating, it is possible to evaluate how this target might affect the navigation process: first of all, VESPA is axis-symmetrical because of its conic shape; the only elements not axis-symmetrical are the superficial elements, such as bolts or flanges. However, those present a central symmetry. This symmetry properties require to arbitrarily define which of the two identical sides of VESPA the chaser faces when the pose is initialized, as the x-y axis orientation is undefined because of the central symmetry. For the tracking, the superficial elements are the only features that could provide information about the chaser position around the axis of symmetry of VESPA.



(a) VESPA detailed model **(b)** VESPA downscaled model and landmarks

Figure 4.2: VESPA models and generated landmarks

4.2 Synthetic images rendering

The visible and thermal images rendering tool was not developed or modified within the context of the thesis but was solely exploited as user. The working principle of the image generation pipeline is briefly presented, and the reader is referred to [45] for a more detailed description. The rendering tool was initially designed for small-bodies rendering and was adapted for the space debris application for the thesis.

Both the VIS and TIR generation relies on Blender rendering software [16]. The process is straightforward for the visible images, as the software is intended for this task. In the case of the thermal images, a simplified thermal model of the target is used to compute the surface irradiance, which is fed into Blender to generate the thermal images. The target simplified thermal model relies on two significant assumptions. Firstly, the target thermal profile is considered spatially uniform at each given instant, so the target appearance in the thermal image is influenced only by the overall temperature, the emissivity of the different surfaces, and the view factor. Secondly, only two thermal conditions were considered: a sunlit (hot) case, which provides a bright thermal image (Figure 4.4b) and an eclipse (cold) case, providing an overall darker image (Figure 4.7b) assuming the target temperature close to lower limit of the sensor's sensibility.

The camera parameters used for the image generation are reported in Table 4.2, where it can be seen that the visible image provides a higher resolution and has a wider FOV with respect to the thermal camera. These parameters are also used to calculate the intrinsic camera matrix to compute the filter's pseudo-measurements.

Parameter	Value	Unit
Resolution	[2000, 2000]	px
FOV	[9.15, 9.15]	deg

(a) VIS camera properties

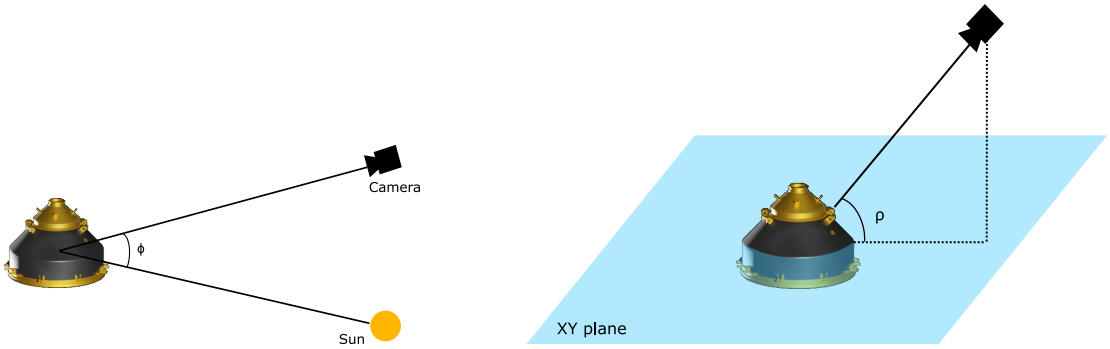
Parameter	Value	Unit
Resolution	[640, 512]	px
FOV	[6.2, 5]	deg

(b) TIR camera properties

Table 4.2: Visible and thermal camera properties

As environmental conditions highly influence space imagery, also the rendered images reflect this behavior. The parameters that most influence the visible and thermal images are identified to be:

- the chaser-target-Sun angle ϕ $[0\ \pi]$ (Figure 4.3a), referred to as *phase angle*, which indicates if the camera is facing the sunlit or shadowed side of the target.
- The *elevation angle* of the chaser in the target body frame ρ $[-\pi/2\ \pi/2]$ (Figure 4.3b), that indicates in which measure the camera observes the outer or inner surface of the conical shape.
- The presence of eclipse.



(a) ϕ angle definition

(b) ρ angle definition

Figure 4.3: Geometrical definition of the phase angles ϕ and the elevation angle ρ

Case	ϕ	ρ	Target visibility		Reference figure
			VIS	TIR	
A	low	high	good	good	Figure 4.4
B	low	low	poor	good	Figure 4.5
C	high	low	poor	good	Figure 4.6
D	eclipse	high	null	poor	Figure 4.7

Table 4.3: Different illumination cases overview

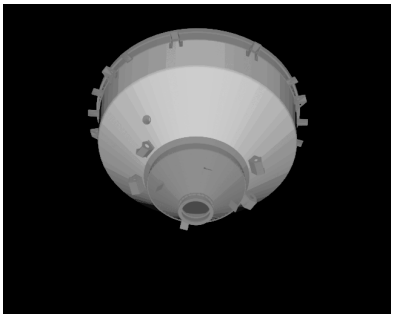
By varying those parameters, four illumination conditions have been identified (Table 4.3). Those study cases solve the role of providing a further description of the synthetic images and provide a better understanding of the filter results in different situations.

- **Case A** represents the most favorable illumination condition for both the visible and thermal spectra. The low phase angle provides good illumination of the target and a high temperature that allows for a vivid thermal image (Figure 4.4b). As highlighted in Figure 4.4a, the high ρ value results in imaging the upper side of VESPA, i.e., the outer side of its conical shape. This represents the best condition for visible imaging, as demonstrated in the second study case.
- **Case B** considers a low phase angle is maintained, but the elevation angle is dropped. It can be observed how observing the concavity of VESPA is critical for the visible imagery (Figure 4.5), as the concavity is subject to important shadows that make the target only partially visible also in good illumination conditions. The thermal image is not highly affected by the shadows, although it shall be noted that the lower side of VESPA has fewer distinguishable elements than its upper side.
- **Case C** considers a situation with a low illumination condition, increasing the phase angle. As expected, even with a high ρ, the visible image is compromised as the target remains only partially visible (Figure 4.6). In this condition, the thermal image is unaffected, as the rendering process works under the assumption of a uniform thermal profile.

- **Case D** investigates the eclipse condition. In that circumstance, the visible image results completely black (Figure 4.6a), as the rendering process assumes the Sun as the only light source. Regarding the thermal images, it is considered the eclipse case as the cold case for the target. The sensor response was lowered to highlight a difference with the hot case, but the temperature of the target is assumed to be still within the sensor sensibility ranges (Figure 4.6b). This was considered because if the target temperature were below the sensor's minimum sensibility also the thermal image would result completely black, becoming a case study not worth investigating as it would produce no results.

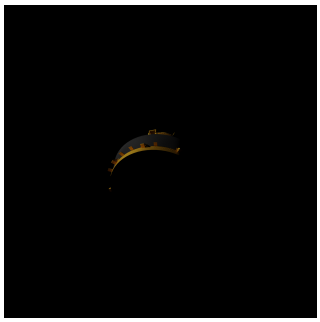
In the thesis work, the images are rendered and used without noise. For further developments, a pixel-level noise should be added to increase the realism of the testing framework [7].

(a) VIS image

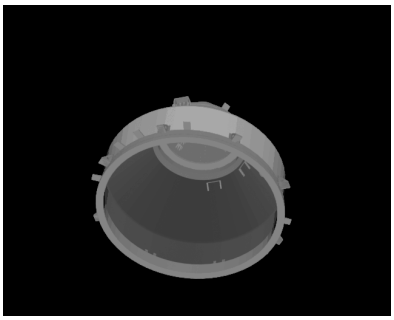


(b) TIR image

Figure 4.4: Case A: low ϕ, high ρ

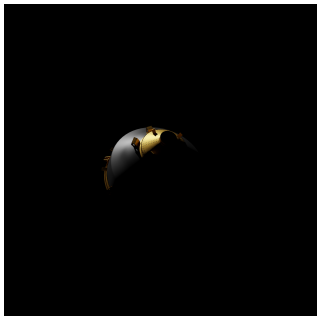


(a) VIS image

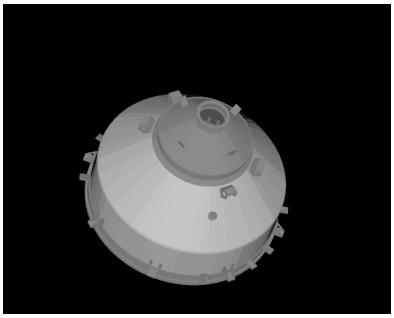


(b) TIR image

Figure 4.5: Case B: low ϕ, low ρ

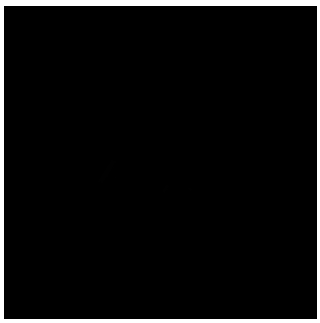


(a) VIS image

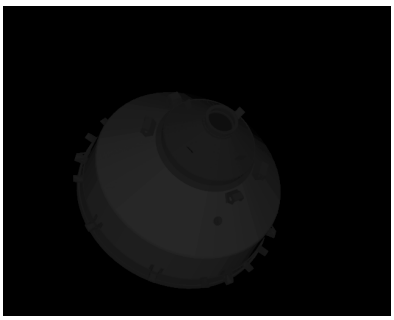


(b) TIR image

Figure 4.6: Case C: high ϕ, high ρ



(a) VIS image



(b) TIR image

Figure 4.7: Case D: eclipse

4.3 Reference dynamics

The Section discusses the reference dynamical model used to test the filter's estimation capabilities.
For what concerns the relative position and velocity, the chaser and the target positions are independently propagated in ECI according to a J2-perturbed Keplerian motion about the Earth. As this model of motion is well known in literature, it is not presented in the thesis. For a detailed description of this dynamics the reader is referred to [3]. The target position is moved into the chaser LVLH frame a posteriori of the propagation. Additional orbital perturbations are not included as their effect is not appreciable in short-term propagations.
As far as the attitude dynamics is concerned, the target angular velocities are propagated with rigid body motion Euler's equation, assuming the principal inertia axes aligned with the body frame axes (Figure 4.2b). The elements of the diagonal inertia matrix were randomly defined as, for the purpose of the thesis, their physical coherence with the target's mass distribution is not relevant. As the chaser mounts two fixed cameras on its body, the cameras' z-axis is assumed to always point the center of the target body. For the sake of simplicity, both cameras are assumed to be placed in the center of the chaser body frame with the cameras' z-axis aligned with the chaser's x-axis (Figure 4.8). This results in the condition that the chaser's x axis shall always point the center of the target's body frame, as if an ideal control was applied. For this reason, the attitude of the chaser is not propagated dynamically, but by ensuring geometrically this condition along the trajectory.

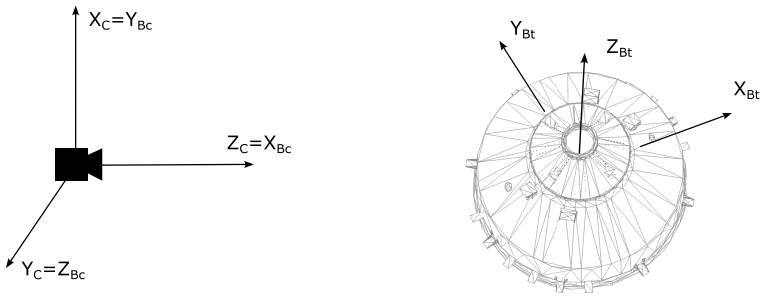


Figure 4.8: Schematic of the chaser, target and camera frames

4.3.1 Figures of merit

After the position and attitude are propagated according to the dynamic described in Section 4.3, and the images are rendered accordingly, the navigation filter is tested to

evaluate its capability to estimate the relative pose. The criteria defined to quantitatively assess the filter's performances are:

- the absolute and relative position error
- the attitude error

This Section defines those variables.
Given a simulation composed on N filter updates (or equivalently N measurements events) the position Absolute Knowledge Error (AKE) at a given step i is defined as:

$$e_{p,i} = \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (z_i - \hat{z}_i)^2} \tag{4.1}$$

where $\hat{\cdot}$ indicates the estimated values. Equation (4.1) can be interpreted as the distance between the true and estimated target position. The position Relative Knowledge Error (RKE) is defined as the absolute error $e_{p,i}$ normalized against the true chaser-target distance:

$$e_{r,i} = \frac{e_{p,i}}{\sqrt{x_i^2 + y_i^2 + z_i^2}} \tag{4.2}$$

It shall be noted that this definition differs from the RKE formulation reported in the ECSS, which defines that index as the difference between the AKE and the mean of the AKE over a time interval.
For the attitude AKE, the Euler angle between the estimated and true relative attitude $e_{a,i}$ is computed. Initially the quaternion error is defined as:

$$q_{err} = q_i \otimes \hat{q}_i' \tag{4.3}$$

where $\hat{q}'$ is the conjugate of the estimated quaternion. The Euler angle associated with the quaternion error can be computed as

$$e_{a,i} = 2\, atan2\left(\sqrt{q_{err,1}^2 + q_{err,2}^2 + q_{err,3}^2}\ ,\ q_{err,4}\right) \tag{4.4}$$

The Euler angle ambiguity between the clockwise or counterclockwise rotation is solved by the $atan2$ operator, which provides always the smallest between the two angles.
The Mean Knowledge Error (MKE) and its associated standard deviation along a simulation are generally defined according to Equations (4.5) and (4.6), where e_i represents a generic error at at the filter step i.

$$\mu_e = \frac{1}{N} \sum_{i=1}^{N} e_i \tag{4.5}$$

$$\sigma_e = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (e_i - \mu_e)^2} \tag{4.6}$$

As the MKE is defined as the average of the error along the simulation, it can be computed for both the AKE and RKE. The mean of the absolute and relative errors will be referred to as Absolute MKE and Relative MKE.

5

Numerical Simulations

The present chapter describes the performances and limitations of the proposed visual navigation pipeline based on the results of numerical simulations.
Using the testing framework described in Chapter 4, the algorithm was stressed in various conditions to understand its advantages and limitations. The testing plan followed is reported in Table 5.1, where the different cases for the illumination conditions are presented in Table 4.3.

Test n.	Illumination condition	Spectra	Scope
1	Case A	VIS & TIR VIS TIR	Evaluate the nominal performances of the multispectral filter, comparing them to the VIS-only and TIR-only applications. (Research questions 1,2)
2	Cases B,C	VIS & TIR	Evaluate the robustness to low illumination for multispectral navigation. (Research questions 1,2)
3	Cases B,D	TIR	Evaluate the robustness of TIR-only navigation under both sunlit and eclipse conditions. (Research question 3)
4	Case A	VIS & TIR	Varying relative distance (20 m to 80 m) to assess the range of applicability of the visual navigation pipeline in terms of chaser-target distance.
5	Case A	VIS & TIR	Synchronous chaser-target rotation to evaluate the influence of apparent dynamics in the visual navigation pipeline.

Table 5.1: Test plan and rationale

At first, the filter's effectiveness is tested in a nominal condition, evaluating the advantages of sensor fusion when both the thermal and visible spectra provide good data quality. Secondly, different aspects of the pipeline are stressed, starting with its dependability from illumination conditions. The possibility of TIR-only navigation is investigated to overcome the aforementioned criticality, presenting its applicability to the test case. Finally, the navigation pipeline robustness is tested against the chaser-target distance and a condition of synchronous rotation.
A comparative assessment is qualitatively performed concerning image fusion techniques applied to navigation about small bodies, as it represents a valuable alternative to the proposed method.

5.1 Test 1: nominal performances

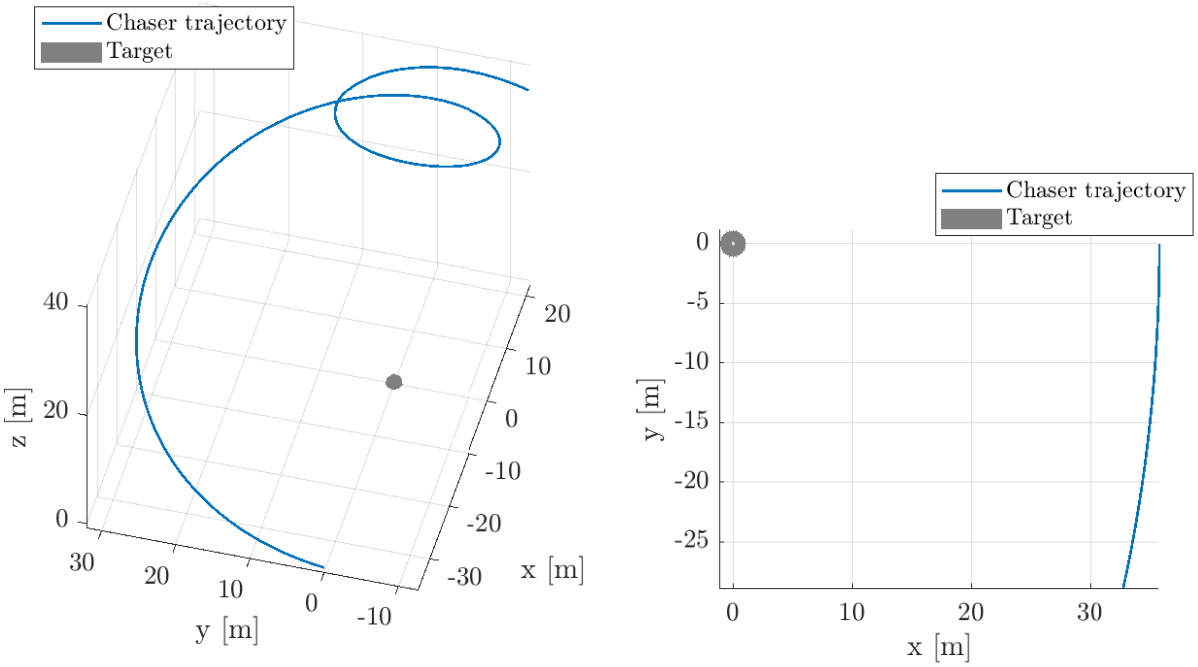
At first, the visual navigation pipeline performances are evaluated in favorable illumination conditions (Case A), having a good target illumination with a phase angle of 25 deg and a distance of 35 m, which do not compromise the Image Processing pipeline as the target is well illuminated and clearly visible on the image plane.
The conditions in which Test 1 is performed are reported in Table 5.2.

ϕ	ρ	Relative distance
25 deg	> 0 deg	35 m

Table 5.2: Test 1 illumination conditions and relative distance

The trajectory followed by the chaser about the target expressed in the target body and LVLH frame is reported in Figure 2.8.

As the thesis work does not include the pose acquisition routine, the initialization of the filter is performed assuming an initial error with respect to the ground truth in terms of position and attitude. The amplitude of the error is selected to be consistent with the results of [43], randomly generating a position offset in the order of 1 m, and an attitude variation of 8 deg maximum. Since the pose acquisition process does not provide information regarding the initial values of the relative velocity or the angular rates, these states are initialized to zero in the filter. The states initialization prior to the addition of the randomic initialization errors is reported in Tables 5.3 and 5.4 for the position and attitude parameters. In Table 5.4 the attitude is parameterized in Euler angles (X ψ, Y ϕ, Z θ) to provide a more straightforward physical interpretation with respect to quaternions.



(a) Target body frame **(b)** Target LVLH frame

Figure 5.1: Chaser trajectory in target's body frame and LVLH frame

x	y	z	$\dot{x}$	$\dot{y}$	$\dot{z}$
35 m	0 m	0 m	$0\,\mathrm{m\,s^{-1}}$	$0\,\mathrm{m\,s^{-1}}$	$0\,\mathrm{m\,s^{-1}}$

Table 5.3: Relative velocity and position initialization

ψ	ϕ	θ	ω_x	ω_y	ω_z
0 deg	0 deg	0 deg	0 deg/s	0 deg/s	0 deg/s

Table 5.4: Relative attitude and angular rates initialization

Both the states' covariance matrix and the process noise covariance matrix are initialized as diagonal matrices. The diagonal elements of the matrix are defined for each state and reported in Table 5.5. As for the covariance matrix, the values are defined to be noticeably higher than the covariance estimate computed by the filter reached steady state. This is performed to account for the initialization error and perform a faster filter convergence to the correct values in the first steps. Considering the filter's dynamic truthfulness, the process noise covariance matrix has been defined with a trial and error procedure to enhance the filter's performance. For the tuning of $\mathbf{Q}$, its influence on the covariance estimation is also considered, trying to avoid either under or over-estimating the states' uncertainty.

Although the measurement noise matrix $\mathbf{R}$ is estimated online by the filter, it is required to provide an initialization as it is estimated recursively. The initial values were

Parameter	Value	Unit
$\sigma^2_{\mathbf{x}}$	2.5e+0	m^2
$\sigma^2_{\dot{\mathbf{x}}}$	8.0e−2	m^2/s^2
$\sigma^2_{\mathbf{a}_p}$	1.0e−2	/
σ^2_{ω}	2.0e−2	rad^2/s^2

(a) Initial covariance matrix settings

Parameter	Value	Unit
$\sigma^2_{\mathbf{x}}$	2.5e+0	m^2
$\sigma^2_{\dot{\mathbf{x}}}$	5.0e−6	m^2/s^2
$\sigma^2_{\mathbf{a}_p}$	3.0e−3	/
σ^2_{ω}	6.0e−4	rad^2/s^2

(b) Process noise covariance diagonal elements

Table 5.5: Diagonal values of the matrices **Q** and **P** at the initial step of the filter

set to be close to the estimates performed by the **R**-adaptation routine to ensure good performances and avoid numerical issues. Similarly to **P** and **Q**, the measurement noise covariance matrix is initialized as a diagonal matrix, whose diagonal elements are reported in Table 5.6. The values are expressed in pixels and refer to non-rectified images, as the measurements are modeled with a pinhole camera model. This is not considered a critical point as the rectification is taken into account by the calibration of the camera in real applications. Given these initialization parameters, the algorithm has been tested on a

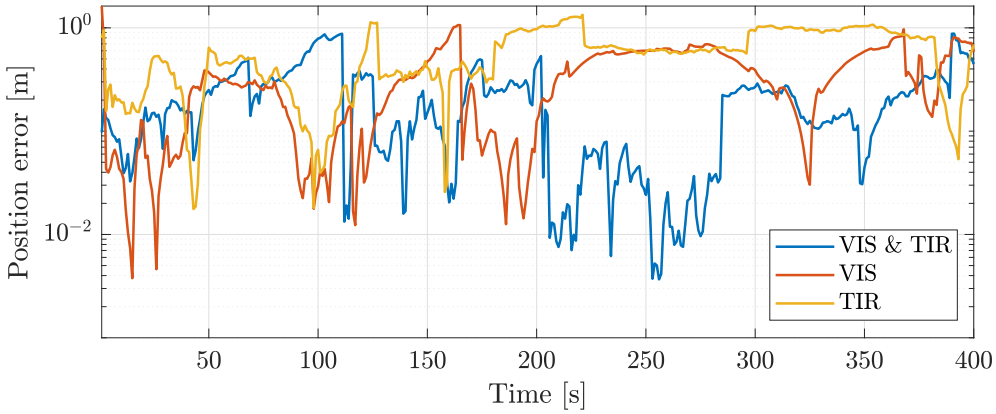
Parameter	Value	Unit
σ_{VIS}	20	px
σ_{TIR}	10	px

Table 5.6: Initial standard deviation assigned to the features

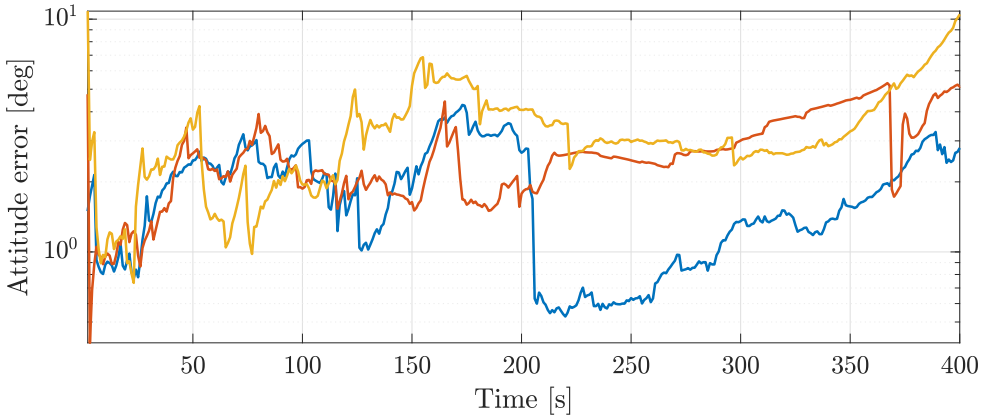
database of four hundred synthetic images generated along the trajectory reported in Figure 2.8. The filter's operating frequency has been set to be 1 Hz. The navigation pipeline has been tested by feeding both the VIS and TIR, the visible only, and the thermal only information in the navigation filter. The results of position and attitude errors for a single run, defined as presented in Section 4.3, are reported in Figures 5.2a and 5.2b respectively.

From Figures 5.2a and 5.2b, it can be qualitatively assessed that the multispectral case can track the position and attitude without ever diverging through the simulation. The multispectral information proved to provide consistently better results than the VIS-only or TIR-only case, as understandable also from the numerical values reported in Table 5.7. This result was expected as in the multispectral case the filter has more information coming from the sensors, thus it can provide a better estimate of the states.

The chaser's estimated trajectory of the chaser in the target body frame, affected by both the position and attitude errors, is represented in Figure 5.3. It can be observed how, after the initial correction of the initialization offset, the estimated path remains bounded to the ground truth on which the sensor's output has been generated.



(a) Position AKE



(b) Attitude AKE

Figure 5.2: Position and attitude errors over a 400 seconds simulation

Spectrum	Absolute MKE [m]	Relative MKE [%]
VIS & TIR	0.25 ± 0.54	0.63 ± 0.54
VIS	0.35 ± 0.59	0.85 ± 0.60
TIR	0.59 ± 0.81	1.51 ± 0.82

(a) Position errors

Spectrum	Absolute MKE [deg]
VIS & TIR	1.72 ± 0.88
VIS	2.64 ± 1.04
TIR	3.36 ± 1.66

(b) Attitude errors

Table 5.7: Position and attitude errors for the different spectrum modalities

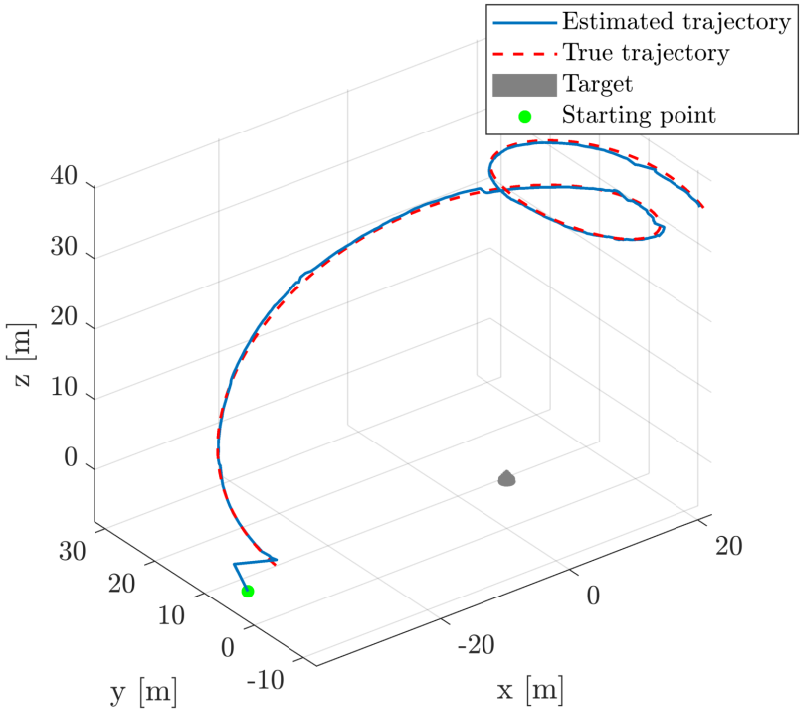
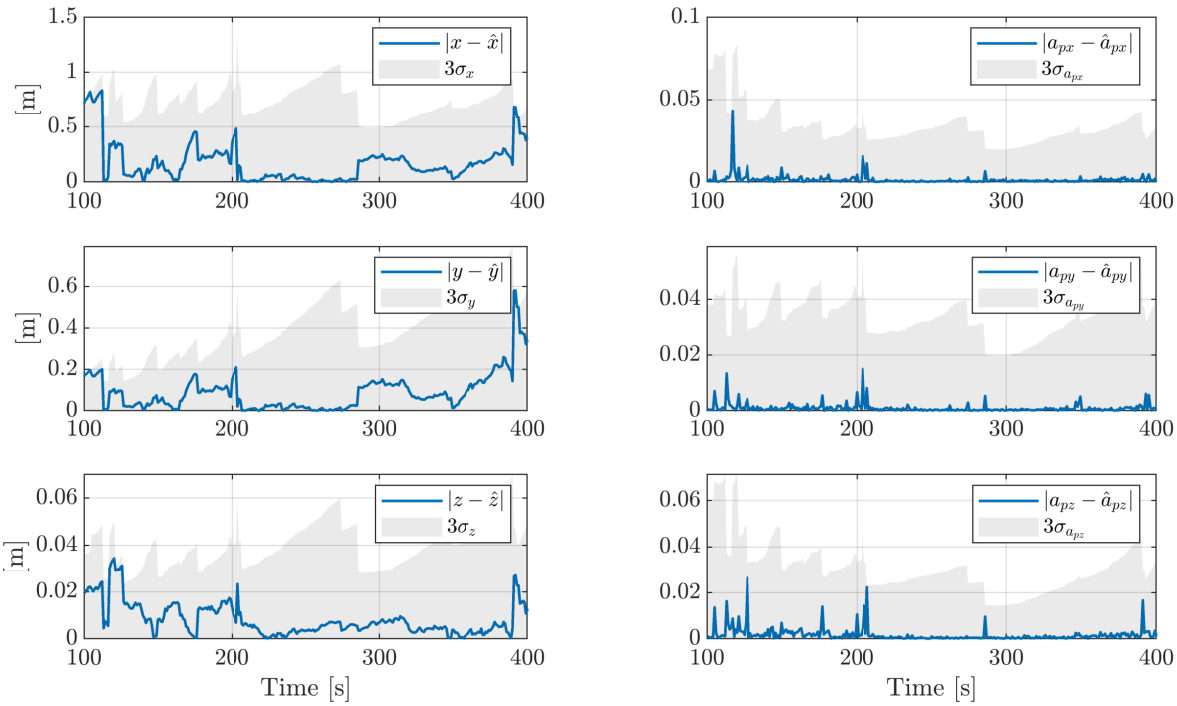


Figure 5.3: Ground truth and estimated trajectory of the chaser about the target body frame

To discuss the confidence in the states' estimation, the position and attitude errors are reported for each component with the associated uncertainty expressed in 3σ. Figure 5.4 shows that the filter provides a reliable estimation of both the relative position and attitude, as the confidence region envelopes the errors without overestimating the uncertainty. The sudden drops in the standard deviation correspond to the re-initialization steps. That is an expected behavior as the uncertainty increase with the decrease in number of features and reduces whenever new features are added.

As the attitude error state expressed in Modified Rodriguez Parameters (Figure 5.4b) does not provide a straightforward physical representation of the problem, the true and estimated Euler angles associated with the relative attitude are reported in Figure 5.5 for completeness. The rotation sequence is X (ψ), Y (ϕ), Z (θ).

The same representation of Figure 5.4 is reported for the relative velocities and angular rates in Figures 5.6a and 5.6b, respectively. The visualization starts at 100 s to enhance the understanding of the graphical representation that would be otherwise compromised by the high error values given by the initialization errors. For both the relative velocities and angular rates, an overestimation of the covariance and an almost null dependence on the feature re-initialization appear evident, contrarily to the relative position or attitude



(a) Position **(b)** Attitude (MRP)

Figure 5.4: Difference between the true and estimated position and attitude (MRP) states with associated uncertainty

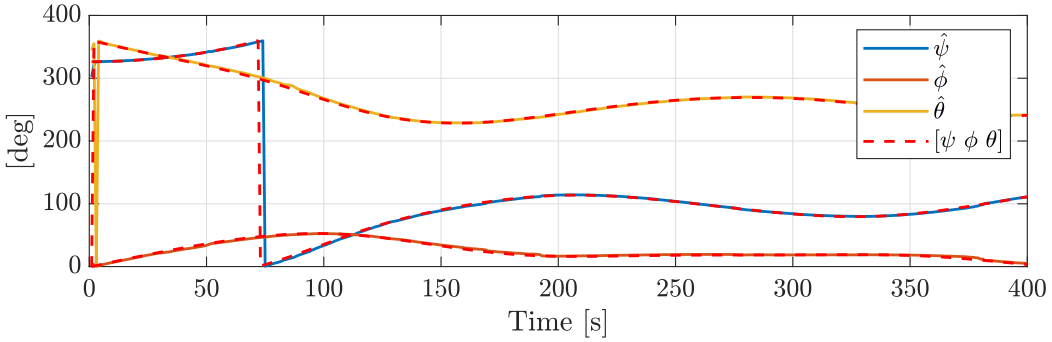


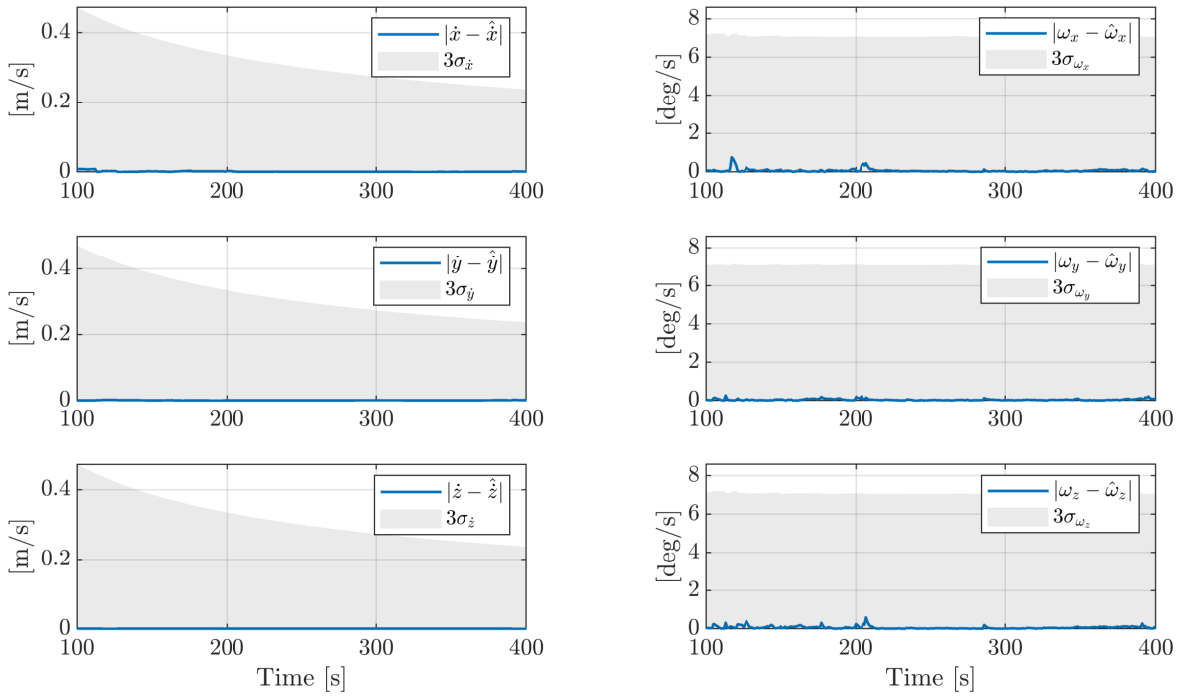
Figure 5.5: True and estimated Euler angles according to XYZ rotation sequence ($\hat{\cdot}$ indicates the estimated values)

case. This latter phenomenon is justified by the fact that there is no direct measurement of the velocity and angular rates.

The uncertainty region associated with the relative velocities decreases as it converges after the initial overestimation due to the initialization error, although presenting a very slow convergence to the steady state value. On the other hand, the overestimation of the relative angular rates results from the increased value of its process noise covariance matrix. Because of the filter's dynamic unreliability, a higher entrustment of the propagated

value would result in a higher error.

The measurement noise matrix adaptation results are reported in Figure 5.7, graphically



(a) Velocity **(b)** Angular rates

Figure 5.6: Difference between the true and estimated relative velocities and angular rates with associated confidence region

illustrating the variance values of the model features at the end of the simulation for both the visible and thermal spectrum. Figure 5.7 shows how the visible and thermal spectra differ in variance values and which features have been matched by the IP routine. As detailed in the following analysis, those differences are one of the reasons behind the TIR spectrum lower performance.

As the filter works at a frequency of 1 Hz, the algorithm must have a computational time smaller than the filter's update time. The computation time of the filter steps and the re-initialization process are reported in Tables 5.8a and 5.8b respectively. Those results are obtained on a Desktop PC with processor Intel Core i5-7300HQ 4 x 2.5 GHz.
The filter run time remains well under one second in all three cases. Since the operations must be performed twice, the multispectral application is almost twice as slow as single-spectrum cases. The re-initialization time is critical, as the computational time is prohibitive as it greatly exceeds the filter's operating frequency of 1 Hz.

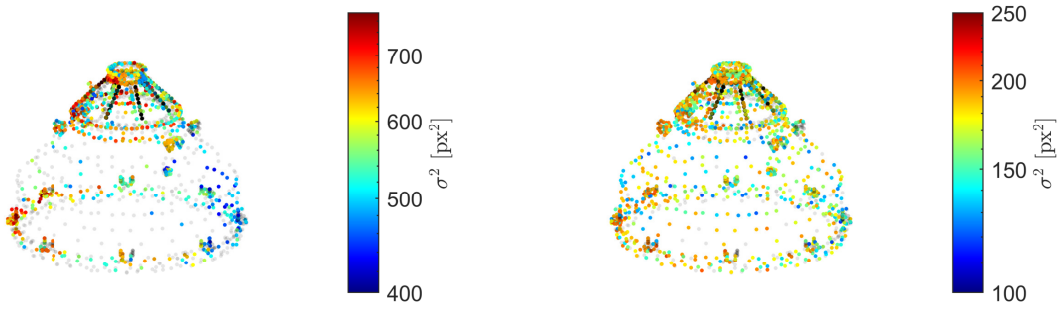


Figure 5.7: Variance of the features at the end of the simulation for VIS (left) and TIR (right) images

The presented times must be reevaluated on relevant hardware to understand the onboard applicability; however, at the current stage, the computational burden is identified as a critical point of the proposed pipeline.

Spectrum	CPU Time [s]
VIS & TIR	0.17 ± 0.03
VIS	0.08 ± 0.02
TIR	0.07 ± 0.02

(a) Filter's step computational time

Spectrum	CPU Time [s]
VIS	8.9 ± 0.55
TIR	9.0 ± 0.46

(b) re-initialization computational time

Table 5.8: CPU time for the the filter step and the reinitialization process

5.2 Navigation pipeline limitations

In the following sections, the visual navigation filter is tested in different conditions (Test 2-5 of Table 5.1) to assess its limitations and range of applicability. The different testing conditions are presented, and the obtained results are discussed to identify the shortcoming and propose possible mitigations.

5.2.1 Test 2: low illumination conditions

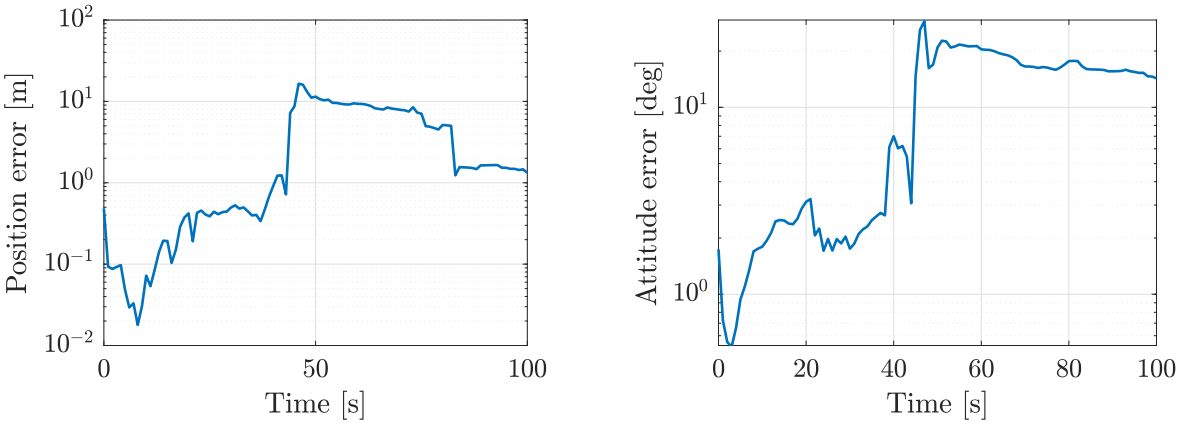
A well-known VIS cameras limitation is the dependence on the target illumination conditions. To assess this condition the algorithm was tested on a database of one-hundred images generated with a phase angle close to 100 deg. In this condition, most of the target is shadowed, and the low quality information coming from the VIS sensors limits the filter performances. A frame of the rendered database is reported in Figure 5.9a, as an example.

The environmental conditions in which the simulation is performed are summarized in Table 5.9.

ϕ	ρ	Relative distance
100 deg	> 0 deg	35 m

Table 5.9: Test 2 illumination conditions and relative distance

The obtained results in terms of position and attitude errors are reported in Figure 5.8: the low illumination conditions strongly affect the results, causing a strictly increasing error in terms of attitude and deteriorating the performance in terms of position estimation. The exceedingly high errors emphasize that the multispectral data-fusion is not suitable for low illumination conditions.



(a) Position AKE

(b) Attitude AKE

Figure 5.8: Position and attitude errors in case of low illumination conditions

The low-quality data provided by the visible sensor have a counterproductive effect on the estimation process, increasing the number of outliers in the measurements and forcing a continuous re-initialization of the features. A small region of the target illuminated decreases the number of features detected in the visible images and concentrates them on a small region of the target, making the pose correction step less effective and affecting it negatively.

As already discussed in Section 4.2, a scarce illumination affecting the visible camera is caused both from an high phase angle (Figure 5.9a) and by a low elevation angle ρ (Figure 5.9b), as the concavity is almost always shadowed.

(a) Sample image in low illumination conditions **(b)** Detail of VESPA concavity

Figure 5.9: Examples of unfavourable visible conditions

This aspect was further investigated highlighting how it affects the feature detection process. Figure 5.10 highlights how the illumination angle is as influential as the chaser elevation angle. It is observed that the only reliable condition is represented by a low phase angle and a high elevation angle.

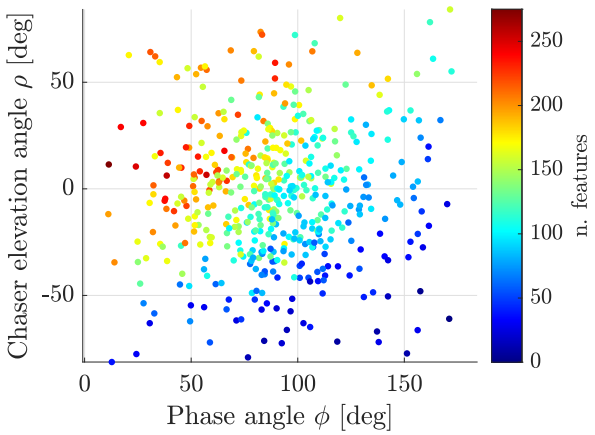


Figure 5.10: Number of matched features as function of the illumination and elevation with respect to VESPA

From Figure 5.9 and the results presented in Figure 5.10, it can be inferred that the multispectral application is limited to a low range of environmental conditions, requiring a good point of view of the target and a good illumination condition.
A possible solution to extend the reliability to illumination conditions was considered to be the application of the Contrast Limited Adaptive Histogram Equalization (CLAHE) to extract information from the shadowed parts as well. However, as presented in Fig-

ure 5.11, in the parts of the image where the target is not illuminated, there is a complete loss of information in terms of pixel intensity, obtaining tiles that are totally black (all pixels have 0 intensity). In this case, the histogram equalization would not enhance but compromise the IP algorithm, adding artifacts in the image. It shall be noted that these considerations apply to those synthetic images, rendered under the assumption of having the Sun as the only light source. Additional illumination sources such as Earth's and Moon's albedo might avoid the presented total black condition, giving the possibility to use CLAHE to enhance the algorithm's performances.
The IP pipeline should be validated on real space imagery to further assess that assertion.

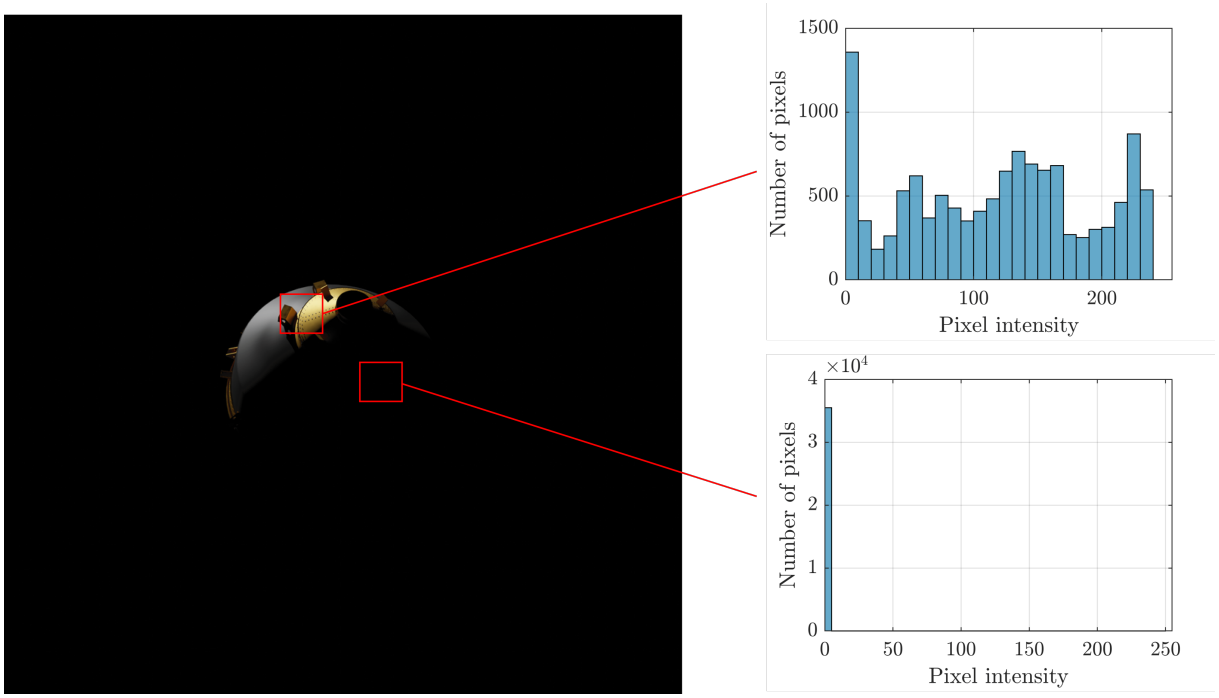


Figure 5.11: Pixel intensity histograms for an illuminated and a shadowed tile in the image.

Since it was assessed that visible sensor exploitation when the target is only partially visible is counterproductive to the estimation process, a possible solution compatible with the proposed pipeline would be to discard the visible sensor output whenever a low illumination condition arises. In this case, it would be necessary for the thermal navigation to provide a reliable pose estimation for long periods of time.
This possibility is investigated in the following analysis.

5.2.2 Test 3: thermal-only navigation

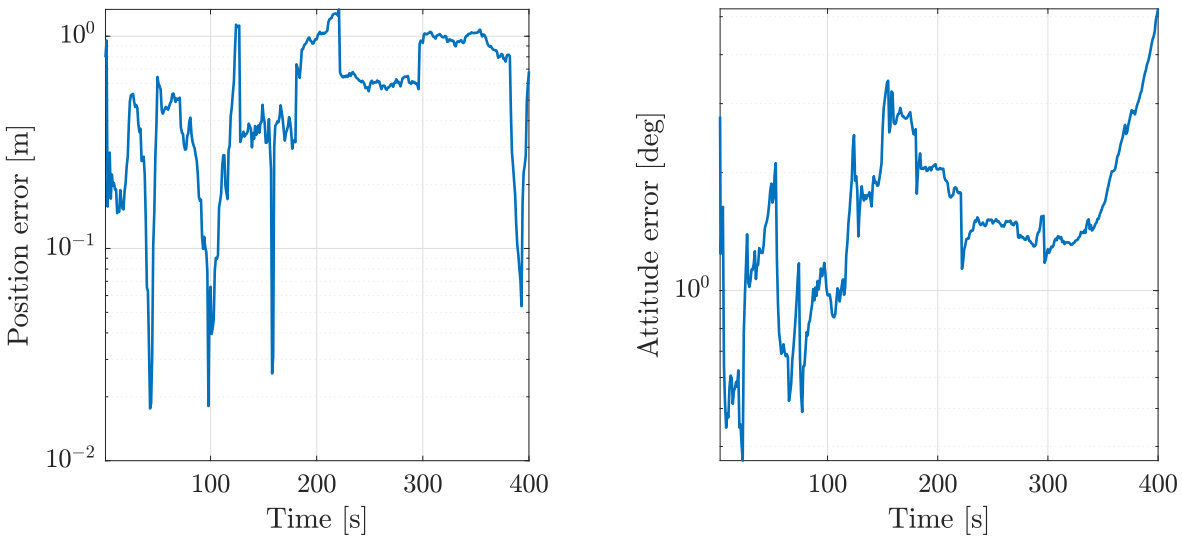
To provide continuous pose estimation in all environmental conditions, the visual navigation filter should function correctly whenever only the thermal images are available. This applies both to conditions of low phase angle, with the target in its hot case, and to eclipse periods, with the target in its cold case.

	ϕ	ρ	Relative distance
Hot case	100 deg	> 0 deg	35 m
Cold case	eclipse	> 0 deg	35 m

Table 5.10: Test 3 illumination conditions and relative distance

Hot case

To test the sunlit condition, the same database generated for Test 1 was used, discarding the visible image. The obtained results in terms of position and angular velocities errors were already shown in Figure 5.2 for comparison with the multispectral case and are reported hereafter (Figure 5.12) for convenience.



(a) Position AKE **(b)** Attitude AKE

Figure 5.12: Position and error for thermal navigation only (hot case)

It can be seen that the position estimation provides stable results, while the attitude counterpart consistently presented an error spike at the end of the simulations, suggesting a future divergence of the filter. To characterize the error behavior's source, the Euler angles between the true and estimated target body frame are computed and presented in

Figure 5.13, with a rotation sequence Z (θ), Y (ϕ), X (ψ).

In Figure 5.13, it can be observed that along the simulation and specifically at the end,

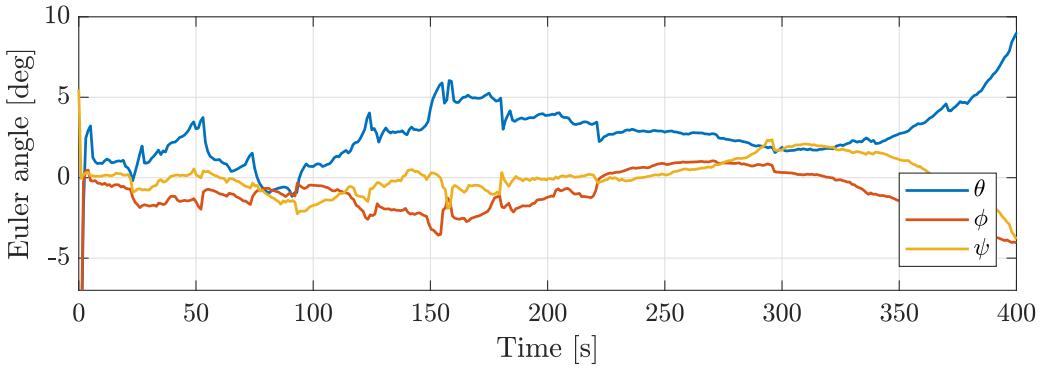


Figure 5.13: Euler angles between the estimated target body frame and the true target body frame (hot case)

the error is mainly concentrated on the z-axis of the VESPA body frame, which is also its axis of symmetry. An error accumulation along this axis indicates that the IP process has difficulties tracking the superficial elements, which, as detailed in Chapter 4, are the only elements providing information about the relative position about the z-axis of the target.

The same consideration could be qualitatively inferred by observing the features tracked by the filter in the first steps of the simulation (Figure 5.14a) and towards the end (Figure 5.14). Both images are subject to CLAHE for contrast enhancement. In the second image, almost no feature is associated with those elements that provide information about the rotation about the axis of symmetry.

This effect is due to the low-quality images in terms of resolution and low contrast, which makes it more difficult to detect the beforementioned elements, differently from the visible case. Moreover, to reduce the rendering tool's computational time, the VESPA model has been down-scaled, splitting the conical shape of the target into a limited number of flat surfaces. That creates an apparent contrast gradient on surfaces that otherwise would be textureless. That contrast discontinuity between these regions allows identifying fictitious features which degrade the Image Processing pipeline as they follow a thermal dynamic instead of the rigid body rotational dynamic of the target, favoring an error about the z-axis.

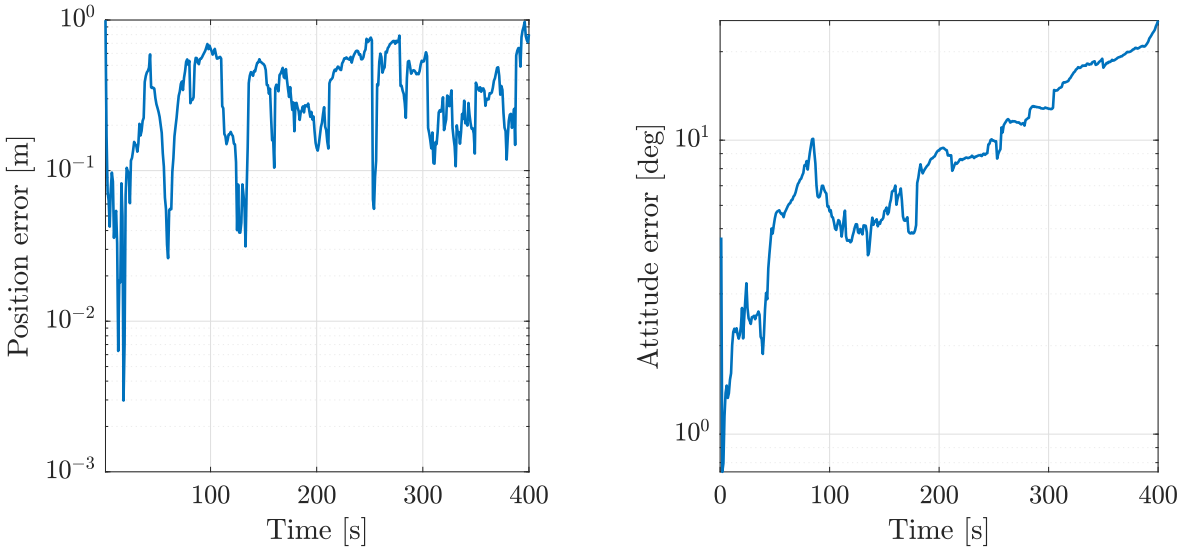
Cold case

Worse results are obtained during the simulated eclipse, as the image quality is lowered with respect to the presented hot case. The position and attitude error results are reported

(a) Beginning of the simulation **(b)** End of the simulation

Figure 5.14: Tracked features on two thermal images of the target

in Figure 5.15. In this case as well, the position estimation works properly, maintaining the error bounded under 1 m. On the contrary, the attitude error presents an almost steadily increasing error, making the estimation diverge.



(a) Position AKE **(b)** Attitude AKE

Figure 5.15: Position and attitude errors for thermal navigation only (cold case)

As for the sunlit case, the Euler angles between the estimated and true target body frame are computed and presented in Figure 5.16. In this case as well, the drift is unique to the z-axis, but more evident than in the sunlit case because of the lower quality of the image.

As shown in Figure 5.17, after applying the histogram equalization to the image in eclipse,

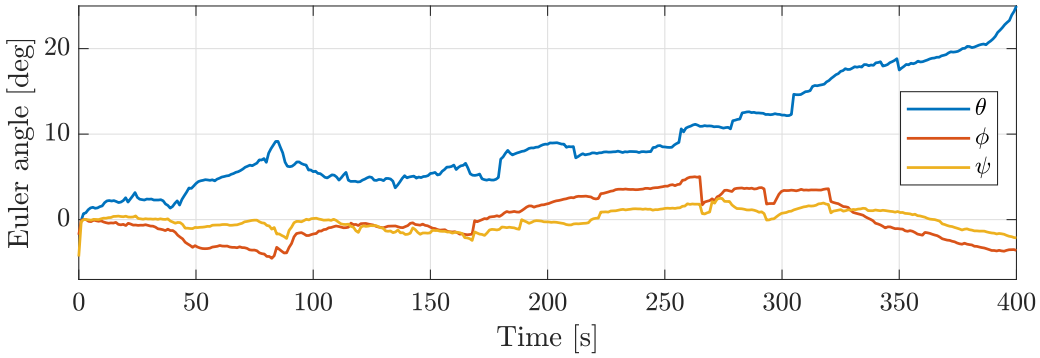


Figure 5.16: Euler angles between the estimated target body frame and the true target body frame (cold case)

a good information on the overall target shape is attainable, while the lack of contrast in the image highly compromises the texture; that highly contributes to the error behavior shown in Figure 5.16, since an error on the x or y axis produces a significant change of the target shape on the image plane, detectable and correctable by the filter, while an error on the z-axis is hardly observable in terms on Image Processing.

(a) Raw TIR image in eclipse

(b) TIR image in eclipse after CLAHE

Figure 5.17: Original and CLAHE enhanced thermal image in eclipse

The obtained results for the sunlit and eclipse case indicate that the proposed navigation filter can not provide reliable standalone pose estimation based on the thermal spectrum only. It shall be noted, however, that most of the criticalities associated with those analyses are related to the symmetrical nature of the target. Further testing on a non-symmetrical target should be performed to test the pipeline in a less critical case, while further work shall be performed to tailor it to apply to symmetrical targets. A possible solution could be to split the IP routine in two segments: one dedicated to the detection of elements that provide information about the relative pose with respect to the x-y axis

of the target and one specifically dedicated to the detection of the superficial elements of VESPA for the estimation of the position and attitude about the z-axis (the axis of symmetry). For the former task it would be interesting to extract the ellipses from the images instead of the point features, as suggested in [34] for navigation about cylindrical spacecrafts.
Finally, the validity of the results proposed is limited by the strong assumptions made on the thermal model of the target. A more refined thermal model should be introduced to further characterize the navigation pipeline to consider aspects such as non-uniform thermal profiles.

5.2.3 Test 4: Target-chaser distance dependability

The visual navigation filter scope is to enhance close proximity navigation capabilities. It is, therefore, necessary to identify the limit of application in terms of distance from the target for which the proposed solution can be used.
Different simulations have been performed on images rendered considering an increasing distance from the target. Because of the time required for image rendering, only four sets of 100 VIS and TIR images have been generated, considering an average distance of 20 m, 40 m, 50 m, and 80 m. Those distances were selected to range from a condition where the target fully occupies the image up to critical distance where the superficial features are hardly identifiable (Figure 5.18). The simulation at different distances have been performed with the same illumination conditions, reported in Table 5.11, to enable a meaningful comparison of the results.

ϕ	ρ	Relative distance
25 deg	> 0 deg	[20 m, 40 m, 50 m, 80 m]

Table 5.11: Test 4 illumination conditions and relative distances

Sample images at different distances are reported in Figures 5.18 and 5.19.

The position Relative MKE and attitude Absolute MKE over a run performed for each distance is reported in Figure 5.20. The numerical results with the associated standard deviation are also reported in Table 5.12. As expected, there is a steady increase in both the parameters estimated by the filter under equal values of ϕ and ρ, as the information provided by the images decreases with the distance.

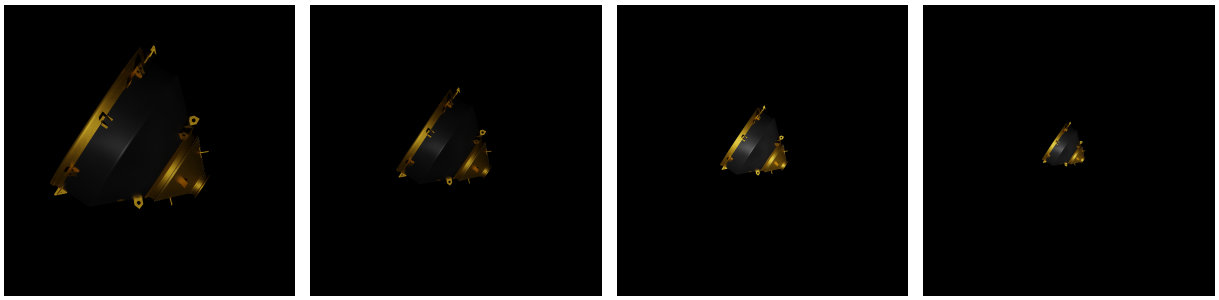


Figure 5.18: Visible images rendered at different distances: 20 m (leftmost), 40 m (left), 50 m (right), 80 m (rightmost)

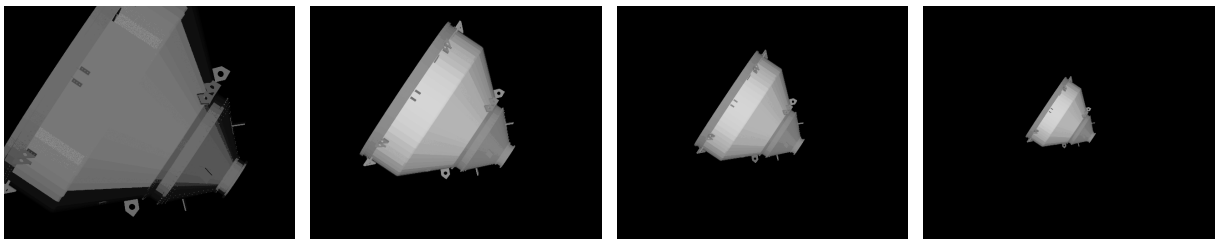


Figure 5.19: Thermal images rendered at different distances: 20 m (leftmost), 40 m (left), 50 m (right), 80 m (rightmost)

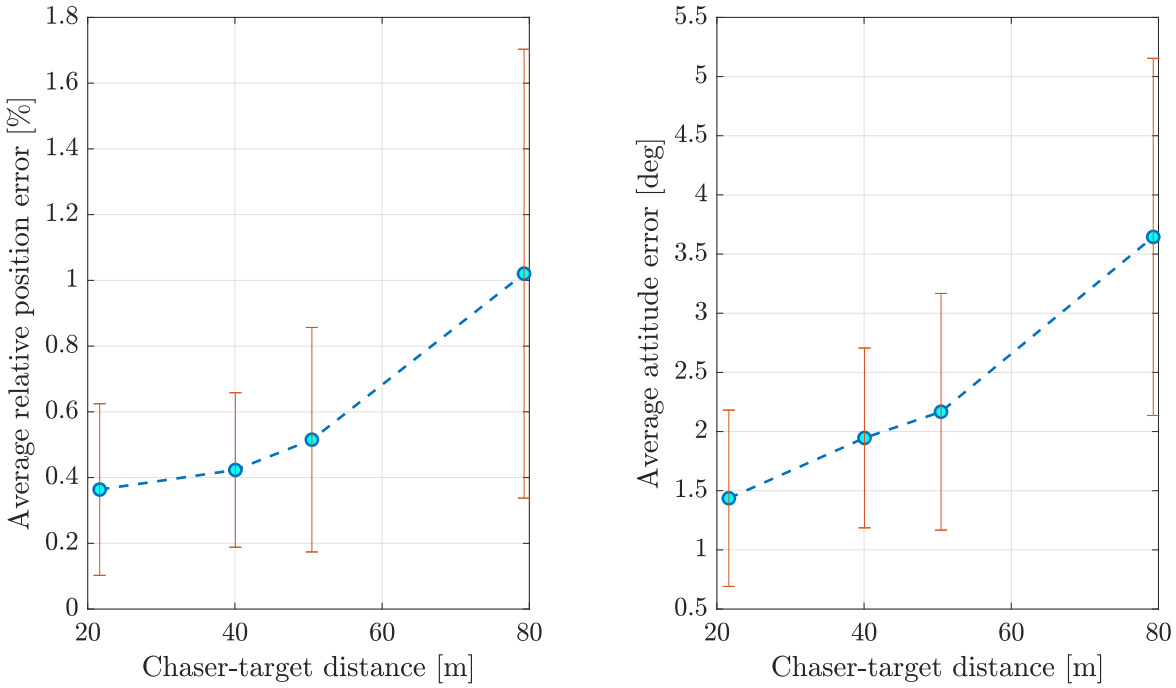


Figure 5.20: Position relative MKE (left) and attitude absolute MKE (right) errors along different simulations with associated standard deviation

Figure 5.20 and Table 5.12 show that with the distance, there is also an increase in the standard deviation of the error. This derives from a high fluctuation in the error values

Distance	Absolute MKE [m]	Relative MKE [%]
20 m	0.07 ± 0.05	0.37 ± 0.26
40 m	0.17 ± 0.10	0.42 ± 0.24
50 m	0.28 ± 0.17	0.52 ± 0.34
80 m	0.82 ± 0.55	1.02 ± 0.68

(a) Position errors

Distance	Absolute MKE [deg]
20 m	1.44 ± 0.72
40 m	1.94 ± 0.76
50 m	2.16 ± 1.00
80 m	3.64 ± 1.50

(b) Attitude errors

Table 5.12: Position and attitude errors at different chaser-target distances

that could bring instabilities in more extended simulations. This performance degradation is attributable to the Image Processing pipeline, as the lower dimension of the target in the image plane compromises the feature detection and matching.
Given the presented results and the cameras' parameters used to render the images, the proposed pipeline is suitable for the close-range inspection and final approach phases, limiting the chaser-target distance to 50 m. To extend the application range, possible solutions include different navigation techniques such as angles-only navigation, either avoiding a 6-DoF pose reconstruction, or including in the sensor suite cameras with different FOVs to provide higher quality images at higher distances.

5.2.4 Test 5: synchronous rotation

A final analysis was carried out investigating the capability of the filter to work correctly in the case there is synchronous rotation between the chaser and the target, giving the appearance that there is no relative motion between the two. This situation is tested in favorable condition for both spectra. The chaser-target distance and the illumination parameters are reported in Table 5.13.

ϕ	ρ	Relative distance
25 deg	≈ 0 deg	35 m

Table 5.13: Test 5 illumination conditions and relative distance

As detailed in the results analysis, two simulations have been performed with the same environmental conditions (Table 5.13), initial conditions and tuning of the filter to highlight the non-deterministic behavior of the navigation filter.
The chaser trajectory in the target body frame can be observed in Figure 5.21. As expected it is almost point-like. This is obtained by setting the angular velocity of the target equal to the opposite of the angular velocity of the chaser position about the target. All other angular velocities are set to zero.

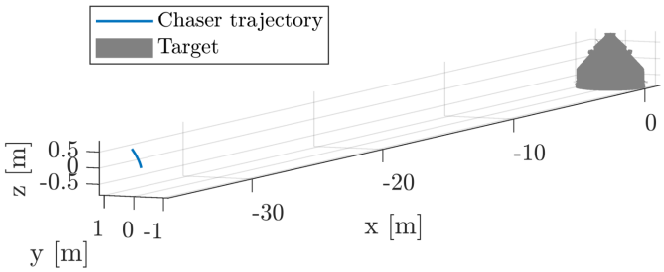
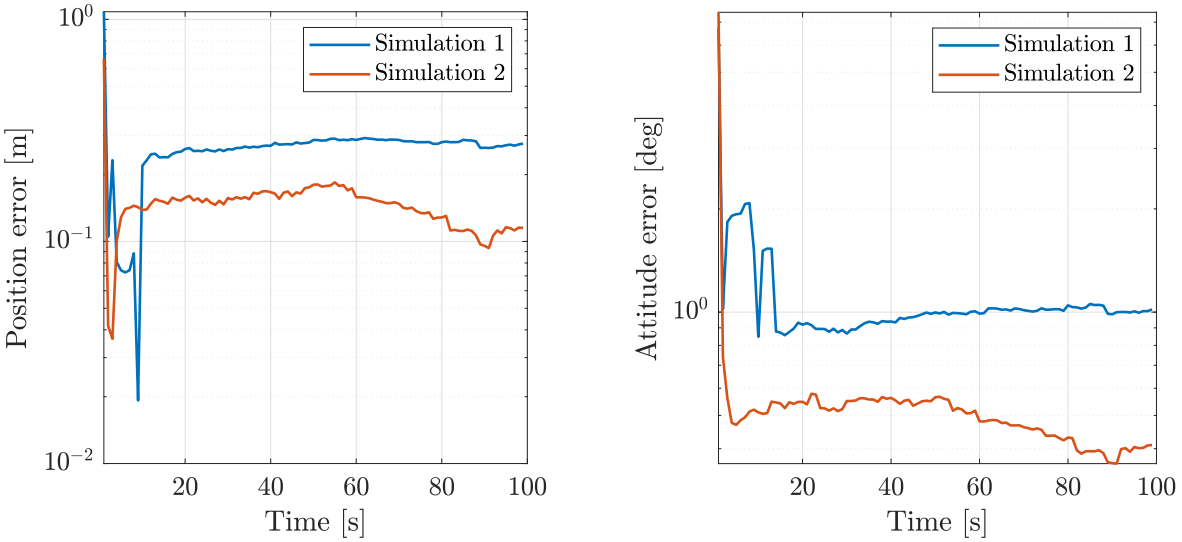


Figure 5.21: Chaser trajectory in the target body frame in the case of quasi-synchronous motion



(a) Position AKE **(b)** Attitude AKE

Figure 5.22: Position and attitude errors for two simulations in the case of quasi-synchronous motion

In Figure 5.22 the error behavior along the two simulations performed can be observed. The two simulations have been executed with an identical initialization of the parameters and environmental conditions; the difference in the results is justified only by the

non-deterministic behavior of the navigation chain. In both cases, the filters present no problems in estimating the attitude and position of the target. Because of the attitude-position dynamical coupling, the apparent stasis of the features on the image plane is recognized as the synchronous rotation of the target with the relative position of chaser. In the case of synchronous rotation, the features are quasi-static on the image plane, requiring for no feature re-initialization as there is no feature loss. This aspect has two main effects: primarily, it increases the filter's performances relaxing the tracking step, resulting in a more stable estimation. Secondly, the filter's performances are mostly influenced by the quality of the first matching between the image points and the model landmarks, which is affected by randomic aspects such as the RANSAC routine. If the association presents an increased number of outliers, the navigation will be consistently affected by an increased error (Simulation 1), while if it provides good feature associations, the errors are lowered (Simulation 2).
From the results presented, the synchronous motion does not represent a critical condition for the proposed navigation pipeline. On the contrary, because of the null relative angular rates, it represents a favorable condition facilitating the IP process.

5.3 Final remarks

The results obtained on the test plan provide an insightful overview of the proposed navigation chain performances and limitations. The key findings for each test case are reported in Table 5.14 to provide a concise yet exhaustive summary of the obtained results.

Test n.	Illumination	Final remarks
1	Case A	Whenever the target is well illuminated and clearly visible both in the VIS and TIR image, the proposed navigation chain can reliably estimate the 6-DoF relative pose. In this condition the fusion of the VIS & TIR information enhances the results, highlighting the positive contribution of sensor fusion. Those results suggest that a tightly coupled filtering approach is suitable for multispectral relative navigation applications. The main limitation identified in this condition is the elevated computational time required for the re-initialization step, which exceeds the filter's update frequency.

2	Cases B,C	The multispectral navigation chain fails if the VIS image is compromised by low illumination conditions, indicating that the visible spectrum is the mainstay of the algorithm's robustness. The phase angle ϕ and the elevation angle ρ resulted to be equally influential on the navigation pipeline performances, restraining the algorithm's application to a limited range of environmental conditions. As the VIS contribution is counterproductive in low illumination conditions, the solution identified is to discard the visible information when this condition arises.
3	Cases B,D	Thermal navigation could not provide reliable pose estimate in both the hot case (Case B) and cold case (Case D). In both cases the navigation chain fails to estimate the relative position about the axis of symmetry of VESPA, resulting in a diverging attitude estimation. The cause of this behavior is identified in the lower performances of the IP routine applied to TIR images, which is not able to track the superficial elements of the target.
4	Case A	The navigation performances are highly influenced by the the chaser-target distance. With the considered camera proprieties it is evaluated that the the proposed application is suited for CPO under a relative distance of 50 m, as it does not provide reliable results if the chaser and the target are further apart.
5	Case A	The navigation chain works properly also under a synchronous motion condition. The dynamical and measurement models enable the filter to estimate all the states correctly also in this peculiar relative motion. Moreover, the performances are improved as the quasi static appearance of the target on the image plane relaxes the IP routine.

Table 5.14: Final remarks for each test case and illumination condition

5.4 Image fusion comparative assessment

In this section, a comparative assessment of the proposed pipeline is performed with respect to the obtained results by [45] regarding multispectral relative navigation performing sensor fusion at image level. The reference work and the thesis investigate the

multispectral navigation problem in two very different situations, as one performs pose estimation about a small body while the other is applied to artificial space objects. Moreover, the asteroid navigation is performed with a SLAM approach, as the target is not initially known, while the thesis work proposes a model-based solution to the navigation problem. Finally, it shall be noted that the analyses performed in [45] are performed with a more realistic testing framework, providing a more reliable insight on the subject. These important differences limit to a qualitative comparison only.

Both methods perform well in good illumination conditions, solving the pose throughout the simulations. For the image fusion technique, it is found that, in nominal conditions, it performs slightly worse than using visible-only or thermal-only images. This is not the case for the proposed pipeline, where the sensor fusion enhances navigation performances. This is due to the introduction of noise in the image fusion process, avoided in the navigation filter. On the other hand, image fusion enables stronger reliability in cases where the target is only partially visible in the VIS image. While the visible contribution in the proposed filter becomes counterproductive, in the image fusion technique, this does not happen, enabling wider multispectral navigation applicability. Finally, in low illumination conditions, both methods are unreliable, showing invalid results.

The consistency of the results in the two applications highlights the need for more maturity of multispectral navigation in scarce illumination conditions, where the thermal sensor should increase the range of applicability. The proposed sensor fusion method outperforms the image-fusion technique when the target is well visible in both the VIS and TIR images. However, the latter is more robust when the visible image is less performing. To the best of the author's knowledge, both methods take advantage of a switch between different modalities based on real-time environmental conditions.

A schematic of the comparison between the proposed filter-based multispectral navigation pipeline and the image fusion-based one proposed in [45] is reported in Table 5.15.

	Filter-based	Image fusion
low ϕ	Reliable pose estimation; VIS & TIR fusion provide the best solution	Reliable pose estimation; VIS only provides the best results
high ϕ	Unreliable pose estimation	Reliable pose estimation; VIS & TIR fusion provides the best results
eclipse/shadow	Unreliable pose estimation	Unreliable pose estimation

Table 5.15: Comparison between the presented filter-based multispectral navigation pipeline and the image fusion based proposed in [45]

Based on the comparison presented in Table 5.15, it appears that image fusion is more advantageous than the method proposed in the thesis because it is more resistant to low illumination of the target. However, due to the significant differences between the two applications and the respective synthetic images, it is not possible to definitively determine which method is superior. To answer the question of whether there is a winner between tightly, loosely, and image-fusion based approaches, it is necessary to apply them to the same test case to have quantitatively comparable results.

6

Conclusion

This thesis presents a study on relative navigation around a known uncooperative target using multispectral imaging sensors. A novel visual navigation pipeline has been developed to exploit images acquired by both a visible and a thermal monocular camera, either by fusing the information or using a single spectrum. The proposed navigation solution has been critically tested on synthetic images of the VESPA debris to evaluate its performance and limitations.

In favorable illumination and thermal conditions, the multispectral data fusion resulted in an enhancement of the pose estimation performance at the expense of a mild increase in computational time. This results demonstrates the proper working of the sensor fusion using a tightly coupled filtering approach applied to the navigation problem, as the EKF successfully fuses the sensor's data providing an improved estimate of the pose. The effectiveness of multispectral data fusion is also enhanced by the adaptivity of the measurement noise matrix, which balances the contribution of the two sensors given their different accuracy.

However, when tested in low illumination conditions, the navigation pipeline was unable to track the relative position and attitude, as the estimated pose consistently diverged from the ground truth. It can be concluded that with the proposed navigation algorithm, the multispectral approach does not provide a reliable standalone solution when different illumination conditions are encountered. The limiting factor is identified as the visible image, which degrades the performance of the Image Processing pipeline. As a consequence, the possibility of exploiting thermal-only navigation when the visible sensor is compromised has been investigated.

The TIR-only navigation application has been tested in both the hot and cold cases of the target. The main difference between the two situations is that in the hot case, the target's temperature is fully within the thermal sensor's temperature range, providing a vivid image. In contrast, in the cold case, the target appears less distinct in the thermal image since its temperature is closer to the lower limit of the sensor's detection range. Although the results are promising, they are not satisfactory enough to confirm thermal-only navigation as a reliable solution. In the hot case, a divergence behavior was identified

at the end of the simulations, while this behavior was prominent from the early stages in the cold case. However, the error drift is mainly caused by the target's symmetry associated with the low performance of Image Processing using only thermal images, which is not able to successfully track the superficial elements of VESPA. This indicates that, with a more distinguishable target shape, thermal navigation might indeed provide robust performances.
From the obtained results, the original research questions can be answered as:

1. Multispectral imaging can improve the relative navigation about a known non-cooperative target. The thermal imaging contribution is limited to an improvement of the performances in good illumination conditions, while it does not successfully increase the range of applications of visible navigation because of the combined degradation of the visible information and the low robustness of the IP routine applied to the thermal images.

2. The most effective imaging measurements management method is to use different spectral modalities in different environmental conditions. If in good illumination conditions the multispectral data fusion increases the navigation performances, in harsher conditions it is beneficial to discard the visible measurements. This highlights an environmental triggered switch between different modalities as the best solution.

3. From the obtained results, thermal sensors can not provide a standalone solution to the relative navigation problem because the low performance of the Image Processing routine applied to thermal images. However, as the limitations are also related to the particular shape of the VESPA target, it is believed that with a more distinguishable target thermal-only navigation might produce valuable results.

4. A tightly coupled sensor fusion approach successfully solves the relative navigation problem. However, because of the dependency of the results on the test case, it is not possible to quantitatively compare the obtained results with other works. It is therefore not identified at which level it is more convenient to fuse the multispectral data.

Even if the validation methodology presents a solid basis, the overall lack of realism in the testing framework may impact the validity of the results. For instance, the absence of noise in the images and background elements creates favorable conditions. However, the simplified structural and thermal model used to generate thermal images introduces fictitious elements not expected in real imagery. To provide a more insightful evaluation of the navigation pipeline, it would be beneficial to design a more realistic testing framework.

6.1 Future work

Because of its relevance for next-generation space missions, autonomous relative navigation about uncooperative target is a continuously evolving subject in the research field. With the growing interest in multispectral navigation in recent years, the obtained results in this thesis work represent a preliminary but promising starting point for further advancement in this area. Some possible future developments identified during the thesis work are here provided:

Testing framework refinement

As mentioned in the previous paragraph, more realistic synthetic images should be implemented to achieve a more insightful assessment of the navigation pipeline's performance. Since both visible and thermal images are currently rendered without noise, it should be added in post-processing, as it may not be negligible for the application considered [7]. An important simplification of the current testing framework is the use of a simplified thermal model for the target. For future studies, it is suggested to increase the realism of the thermal model by considering transient behaviors and a non-uniform temperature profile of the target. In this context, the work proposed by [13] presents a refined thermal rendering tool tailored for space imagery.

Multispectral navigation

In this work, the data fusion has been performed at feature level, demonstrating the possibility to perform multispectral data fusion with a tightly coupled approach. However, the results suggest that the proposed multispectral algorithm provides a meaningful solution only when the target is clearly visible in both thermal and visible images. As one proposed approach to tackle this issue involves using thermal navigation as a standalone solution for low illumination conditions, further effort should be focused on making this method more reliable. One advantage of the proposed solution is its flexibility to changes in the Image Processing routine, enabling the investigation of hybrid approaches for the visible and thermal spectrum. An IP routine tailored specifically for thermal images or using more innovative techniques, such as CNNs, could provide an important contribution to the applicability of the proposed navigation filter.

Extension to different mission scenarios

The visual navigation pipeline developed has been presented and tested in an ADR scenario. However, it can be adjusted for other space missions involving navigation around a known target, such as OOS or FF. As the navigation filter provided promising results for a peculiar target such as VESPA, it is expected that testing the pipeline on targets with more distinguishable shapes will yield even better results than those presented in the thesis. Moreover, the flexibility of the tightly coupled approach allows for straightforward integration of additional instruments in the sensor suite, making it easy to adapt for cooperative relative navigation.
The navigation solution could also be adapted for small-body exploration, assuming either a known asteroid or successful mapping of it. In this context, it would be interesting to compare this work with that proposed in [12], where a loosely coupled approach was used to navigate around the Ryugu asteroid.

Bibliography

[1] European Space Agency. *VV02 – Vega uses Vespa*. https://www.esa.int/Enabling_Support/Space_Transportation/Launch_vehicles/VV02_Vega_uses_Vespa. Accessed: 30/03/2022.

[2] Shahrokh Akhlaghi, Ning Zhou, and Zhenyu Huang. "Adaptive adjustment of noise covariance in Kalman filter for dynamic state estimation". In: *2017 IEEE power & energy society general meeting*. IEEE. 2017, pp. 1–5.

[3] Kyle T Alfriend et al. *Spacecraft formation flying: Dynamics, control and navigation*. Vol. 2. Elsevier, 2009.

[4] Simon Baker and Iain Matthews. "Lucas-kanade 20 years on: A unifying framework". In: *International journal of computer vision* 56 (2004), pp. 221–255.

[5] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features". In: *Lecture notes in computer science* 3951 (2006), pp. 404–417.

[6] M Bechini et al. "Robust Monocular Pose Initialization via Visual and Thermal Image Fusion". In: *INTERNATIONAL ASTRONAUTICAL CONGRESS: IAC PROCEEDINGS*. 2022, pp. 1–15.

[7] Michele Bechini, Michèle Lavagna, and Paolo Lunghi. "Dataset generation and validation for spacecraft pose estimation via monocular images processing". In: *Acta Astronautica* 204 (2023), pp. 358–369. ISSN: 0094-5765. DOI: https://doi.org/10.1016/j.actaastro.2023.01.012. URL: https://www.sciencedirect.com/science/article/pii/S0094576523000127.

[8] Robin Biesbroek et al. "The clearspace-1 mission: Esa and clearspace team up to remove debris". In: *Proc. 8th Eur. Conf. Sp. Debris*. 2021, pp. 1–3.

[9] Jean-Yves Bouguet et al. "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm". In: *Intel corporation* 5.1-10 (2001), p. 4.

[10] Lorenzo Pasqualetto Cassinis, Robert Fonod, and Eberhard Gill. "Review of the robustness and applicability of monocular pose estimation systems for relative navigation with an uncooperative spacecraft". In: *Progress in Aerospace Sciences* 110 (2019), p. 100548.

[11] Bruno Cavrois et al. "LIRIS demonstrator on ATV5: a step beyond for European non cooperative navigation system". In: *AIAA guidance, navigation, and control conference*. 2015, p. 0336.

[12] Gaia Letizia Civardi, M Piccinin, and M Lavagna. "Small bodies IR imaging for vision based relative navigation and mapping enhancement". In: *7th IAA Planetary Defense Conference*. 2021, p. 109.

[13] Gaia Letizia Civardi et al. "Generation of fused visible and thermal-infrared images for uncooperative spacecraft proximity navigation". In: *Advances in Space Research* (2023). ISSN: 0273-1177. DOI: https://doi.org/10.1016/j.asr.2023.03.022. URL: https://www.sciencedirect.com/science/article/pii/S0273117723002144.

[14] Gaia Letizia Civardi et al. "Vis-tir imaging for uncooperative objects proximity navigation: a tool for development and testing". In: *11th International Workshop on Satellite Constellations Formation Flying (IWSCFF 2022), International Astronautical Federation, IAF, Milan, Italy*. Vol. 6. 2022.

[15] WH Clohessy and RS Wiltshire. "Terminal guidance system for satellite rendezvous". In: *Journal of the Aerospace Sciences* 27.9 (1960), pp. 653–658.

[16] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation. Stichting Blender Foundation, Amsterdam, 2018. URL: http://www.blender.org.

[17] Philip David et al. "SoftPOSIT: Simultaneous pose and correspondence determination". In: *International Journal of Computer Vision* 59 (2004), pp. 259–284.

[18] Elan Dubrofsky. "Homography estimation". In: *Diplomová práce. Vancouver: Univerzita Britské Kolumbie* 5 (2009).

[19] Hugh Durrant-Whyte and Tim Bailey. "Simultaneous localization and mapping: part I". In: *IEEE robotics & automation magazine* 13.2 (2006), pp. 99–110.

[20] Martin A Fischler and Robert C Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography". In: *Communications of the ACM* 24.6 (1981), pp. 381–395.

[21] Gabriella Gaias and Jean-Sébastien Ardaens. "Flight Demonstration of Autonomous Noncooperative Rendezvous in Low Earth Orbit". In: *Journal of Guidance, Control, and Dynamics* 41.6 (2018), pp. 1337–1354. DOI: 10.2514/1.G003239.

[22] Gabriella Gaias and Jean-Sébastien Ardaens. "Flight demonstration of autonomous noncooperative rendezvous in low earth orbit". In: *Journal of Guidance, Control, and Dynamics* 41.6 (2018), pp. 1337–1354.

[23] Joseph M Galante et al. "Fast Kalman filtering for relative spacecraft position and attitude estimation for the raven ISS hosted payload". In: *AAS Guidance and Control Conference*. GSFC-E-DAA-TN29232. 2016.

[24] Scott P Geyer et al. "Relative Navigation of Uncooperative Space Bodies". In: *AIAA Scitech 2021 Forum*. 2021, p. 1865.

[25] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[26] *Space systems — Space debris mitigation requirements*. Standard. Geneva, CH: International Organization for Standardization, July 2019.

[27] Yupeng Jia et al. "Lvio-fusion: A self-adaptive multi-sensor fusion slam framework using actor-critic method". In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2021, pp. 286–293.

[28] Thomas Karlsson et al. "Prisma irides: Performance at the end of the drift phase and planned rendezvous experiments". In: *Proceedings of the 9th International ESA Conference on Guidance, Navigation & Control Systems, European Space Agency, ESTEC, Noordwijk, The Netherlands*. 2014.

[29] Yu Dam Lee, La Woo Kim, and Hyung Keun Lee. "A tightly-coupled compressed-state constraint Kalman Filter for integrated visual-inertial-Global Navigation Satellite System navigation in GNSS-Degraded environments". In: *IET Radar, Sonar & Navigation* 16.8 (2022), pp. 1344–1363.

[30] Ern J Lefferts, F Landis Markley, and Malcolm D Shuster. "Kalman filtering for spacecraft attitude estimation". In: *Journal of Guidance, control, and Dynamics* 5.5 (1982), pp. 417–429.

[31] Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua. "EP n P: An accurate O (n) solution to the P n P problem". In: *International journal of computer vision* 81 (2009), pp. 155–166.

[32] Stefan Leutenegger, Margarita Chli, and Roland Y Siegwart. "BRISK: Binary robust invariant scalable keypoints". In: *2011 International conference on computer vision*. Ieee. 2011, pp. 2548–2555.

[33] JC Liou. "Active debris removal and the challenges for environment remediation". In: *28th International Symposium on Space Technology*. JSC-CN-23921. 2012.

[34] Chang Liu and Weiduo Hu. "Relative pose estimation for cylinder-shaped spacecrafts using single image". In: *IEEE Transactions on Aerospace and Electronic Systems* 50.4 (2014), pp. 3036–3056.

[35] G Lowe. "Sift-the scale invariant feature transform". In: *Int. J* 2.91-110 (2004), p. 2.

[36] Bruce D Lucas and Takeo Kanade. "An iterative image registration technique with an application to stereo vision". In: *IJCAI'81: 7th international joint conference on Artificial intelligence*. Vol. 2. 1981, pp. 674–679.

[37] F Landis Markley and John L Crassidis. *Fundamentals of spacecraft attitude determination and control*. Vol. 1286. Springer, 2014.

[38] Patrick Michel et al. "The ESA Hera mission: detailed characterization of the DART impact outcome and of the binary asteroid (65803) Didymos". In: *The Planetary Science Journal* 3.7 (2022), p. 160.

[39] Sherif AS Mohamed et al. "A survey on odometry for autonomous navigation systems". In: *IEEE access* 7 (2019), pp. 97466–97486.

[40] ESA Space Debris Office. *ESA's annual space environment report*. Tech. rep. Technical Report GEN-DB-LOG-00288-OPS-SD, ESA Space Debris Office, 2022.

[41] Tatsuaki Okada et al. "Thermal infrared imaging experiments of C-type asteroid 162173 Ryugu on Hayabusa2". In: *Space Science Reviews* 208 (2017), pp. 255–286.

[42] Roberto Opromolla et al. "A review of cooperative and uncooperative spacecraft pose determination techniques for close-proximity operations". In: *Progress in Aerospace Sciences* 93 (2017), pp. 53–72.

[43] Vincenzo Pesce et al. "Autonomous relative navigation around uncooperative spacecraft based on a single camera". In: *Aerospace Science and Technology* 84 (2019), pp. 1070–1080.

[44] Vincenzo Pesce et al. "Comparison of filtering techniques for relative attitude estimation of uncooperative space objects". In: *Aerospace Science and Technology* 84 (2019), pp. 318–328.

[45] Margherita Piccinin. "Spacecraft relative navigation with electro-optical sensors around uncooperative targets". PhD thesis. 2023.

[46] Margherita Piccinin et al. "Multispectral Imaging Sensors for asteroids relative navigation". In: *INTERNATIONAL ASTRONAUTICAL CONGRESS: IAC PROCEEDINGS*. 2021, pp. 1–12.

[47] Robin Pinson, Richard Howard, and Andrew Heaton. "Orbital express advanced video guidance sensor: ground testing, flight results and comparisons". In: *AIAA guidance, navigation and control conference and exhibit*. 2008, p. 7318.

[48] Duarte Rondao and Nabil Aouf. "Multi-view monocular pose estimation for spacecraft relative navigation". In: *2018 AIAA Guidance, Navigation, and Control Conference*. 2018, p. 2100.

[49] Duarte Rondao et al. "Benchmarking of local feature detectors and descriptors for multispectral relative navigation in space". In: *Acta Astronautica* 172 (2020), pp. 100–122.

[50] Edward Rosten and Tom Drummond. "Fusing points and lines for high performance tracking". In: *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. Vol. 2. Ieee. 2005, pp. 1508–1515.

[51] Ethan Rublee et al. "ORB: An efficient alternative to SIFT or SURF". In: *2011 International conference on computer vision*. Ieee. 2011, pp. 2564–2571.

[52] Sumant Sharma, Connor Beierle, and Simone D'Amico. "Pose estimation for non-cooperative spacecraft rendezvous using convolutional neural networks". In: *2018 IEEE Aerospace Conference*. IEEE. 2018, pp. 1–12.

[53] Sumant Sharma and Simone D'Amico. "Reduced-dynamics pose estimation for non-cooperative spacecraft rendezvous using monocular vision". In: *38th AAS Guidance and Control Conference, Breckenridge, Colorado*. Vol. 2. 2017.

[54] J Shi, Steve Ulrich, and Stéphane Ruel. "Spacecraft pose estimation using a monocular camera". In: *67th International Astronautical Congress*. Guadalajara. 2016.

[55] Jian-Feng Shi et al. "Uncooperative spacecraft pose estimation using an infrared camera during proximity operations". In: *AIAA SPACE 2015 conference and exposition*. 2015, p. 4429.

[56] Simrandeep Singh, Nitin Mittal, and Harbinder Singh. "Review of various image fusion algorithms and image fusion performance metric". In: *Archives of Computational Methods in Engineering* 28 (2021), pp. 3645–3659.

[57] A Sonnenburg. "Image recognition and processing for navigation (irpn)". In: *ESA Clean Space Industrial Days, ESTEC, Netherlands* (2016).

[58] Joshua Sullivan, Sebastian Grimberg, and Simone D'Amico. "Comprehensive survey and assessment of spacecraft relative motion dynamics models". In: *Journal of Guidance, Control, and Dynamics* (2017).

[59] Brent E Tweddle and Alvar Saenz-Otero. "Relative computer vision-based navigation for small inspection spacecraft". In: *Journal of guidance, control, and dynamics* 38.5 (2015), pp. 969–978.

[60] Renato Volpe, Marco Sabatini, and Giovanni B Palmerini. "Pose and shape reconstruction of a noncooperative spacecraft using camera and range measurements". In: *International Journal of Aerospace Engineering* 2017 (2017).

[61] M Yakovlev. "The "iadc space debris mitigation guidelines" and supporting documents". In: *4th European Conference on Space Debris*. Vol. 587. ESA Publications Division Noordwijk, The Netherlands. 2005, pp. 591–597.

[62] Koji Yamanaka and Finn Ankersen. "New state transition matrix for relative motion on an arbitrary elliptical orbit". In: *Journal of guidance, control, and dynamics* 25.1 (2002), pp. 60–66.

[63] Mengshen Yang et al. "Sensors and sensor fusion methodologies for indoor odometry: A review". In: *Polymers* 14.10 (2022), p. 2019.

[64] Ozgun Bursalioglu Yilmaz et al. "Using infrared based relative navigation for active debris removal". In: 2017.

[65] Özgün Yılmaz et al. "Thermal analysis of space debris for infrared-based active debris removal". In: *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering* 233.3 (2019), pp. 811–822.

[66] Paul Zarchan. *Progress in astronautics and aeronautics: fundamentals of Kalman filtering: a practical approach*. Vol. 208. Aiaa, 2005.

[67] Lijun Zhang et al. "Unscented Kalman filtering for relative spacecraft attitude and position estimation". In: *The Journal of Navigation* 68.3 (2015), pp. 528–548.

[68] Karel Zuiderveld. "Contrast limited adaptive histogram equalization". In: *Graphics gems* (1994), pp. 474–485.