# Monocular Vision-Based Pose Estimation of Uncooperative Spacecraft

Pasqualetto Cassinis, L.

**DOI**
10.4233/uuid:27dcbbc2-7d9e-4f67-925a-5e676ca4e43c

**Publication date**
2022

**Document Version**
Final published version

**Citation (APA)**
Pasqualetto Cassinis, L. (2022). *Monocular Vision-Based Pose Estimation of Uncooperative Spacecraft*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:27dcbbc2-7d9e-4f67-925a-5e676ca4e43c

**Important note**
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

# MONOCULAR VISION-BASED POSE ESTIMATION OF UNCOOPERATIVE SPACECRAFT

# MONOCULAR VISION-BASED POSE ESTIMATION OF UNCOOPERATIVE SPACECRAFT

## Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op woensdag 16 november 2022 om 15:00 uur

door


**LORENZO PASQUALETTO CASSINIS**

Ingenieur Luchtvaart- en Ruimtevaart,
Technische Universiteit, Delft, Nederland,
geboren te Venetië, Italië.

Dit proefschrift is goedgekeurd door de promotoren. Samenstelling promotiecommissie:

| | |
|---|---|
| Rector Magnificus, | voorzitter |
| Prof. dr. E.K.A. Gill, | Technische Universiteit Delft, promotor |
| Dr. A. Menicucci, | Technische Universiteit Delft, promotor |
| | |
| *Onafhankelijke leden:* | |
| Prof. dr. P. Visser, | Technische Universiteit Delft |
| Prof. dr. S. D'Amico, | Stanford University |
| Prof. dr. O. Deniz, | Universidad de Castilla-La Mancha |
| Prof. dr. M. Lavagna, | Politecnico di Milano |
| Dr. H. Frei, | Deutsches Zentrum für Luft- und Raumfahrt |
| Prof. dr. M. Mulder, | Technische Universiteit Delft, reservelid |





*Keywords:* Relative Pose Estimation, Active Debris Removal, Relative Navigation, On-ground Validation, Artificial Intelligence
*Printed by:* Ipskamp printing
*Front & Back:* Lorenzo Pasqualetto Cassinis



ISBN 000-00-0000-000-0

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

*Wayfarer, there is no path,*
*you make the path as you go.*

Antonio Machado

# CONTENTS

# SUMMARY

Activities in outer space have entered a new era of growth, fostering human development and improving key Earth-based applications such as remote sensing, navigation, and telecommunication. The recent creation of SpaceX's Starlink constellation as well as the steep increase in CubeSat launches are expected to revolutionize the way we use space and extend the current capabilities of satellite-based technology. However, this steep increase in the number of human-made objects is rapidly leading to higher collision risks in congested Earth orbits. This has led to questioning whether this trend is sustainable on the long term, and ultimately to the need to tackle sustainability in space.

The recent decade has seen considerable efforts by Space Agencies to both prevent major collisions in orbit via Active Debris Removal (ADR) missions and to extend the lifetime of the functioning satellites with On-Orbit Servicing (OOS). Unfortunately, the approach and capture of space debris objects is complicated by the fact that these targets are uncooperative and cannot aid close-proximity operations, leading to critical challenges in the estimation of their relative position and attitude (pose) with respect to the servicer spacecraft. Several missions have been proposed as technology demonstrators of debris removal and servicing technologies, in which passive monocular cameras are combined with active sensors to improve the robustness and accuracy of the navigation system. Yet, despite the inherent challenges that come together with the use of monocular cameras in space, navigation systems based on a single camera are becoming an attractive alternative to systems based on active sensors, due to their reduced mass, power consumption and system complexity. The research work presented in this thesis aims at developing and validating a robust and accurate monocular camera-based pose estimation system compliant with navigation requirements of both ADR and OOS missions.
Two fundamental open challenges are addressed:

1. The robustness and applicability of image processing algorithms and pose estimation methods.

2. The validation of relative navigation filters and their interface with image processing and pose estimation.

This research begins with a survey on the robustness and applicability of existing monocular vision-based pose estimation systems. After identifying the characteristics and limitations of each subsystem implemented in state-of-the-art architectures, a comparative assessment of the current solutions is given at different levels of the pose estimation process, in order to bring a novel and broad perspective. Special focus is put on the improved robustness of novel image processing schemes and pose estimators based on Convolutional Neural Networks (CNN). The limitations and drawbacks of the validation of current pose estimation schemes with synthetic images are further discussed, together with the critical trade-offs for the selection of visual-based navigation filters.

Building on the results of the survey, a novel framework is introduced to enable a robust and accurate pose estimation. Two investigated CNNs are used at image processing level to identify a set of pre-selected features on the target spacecraft, which are fed to a pose estimator prior to the navigation filter (loosely-coupled) or directly to the navigation filter as measurements (tightly-coupled). A novel method to derive covariance matrices directly from the CNN heatmaps is introduced to improve the modeling of the feature detection uncertainty prior to pose estimation. The performance results indicate that a tightly-coupled approach can guarantee an advantageous coupling between the rotational and translational states within the filter, while reflecting a representative measurements covariance. Synthetic monocular images of the European Space Agency's Envisat spacecraft are used to generate datasets for training, validation and testing of the CNN. Likewise, the images are used to recreate a representative close-proximity scenario for the validation of the proposed filter.

This research work then extends the validation from a purely synthetic one to a more comprehensive on-ground validation. To this end, ESA's GNC Rendezvous, Approach and Landing Simulator testbed is used to validate the proposed CNN-based pose estimation system on representative rendezvous scenarios, with special focus on solving the domain shift problem which characterizes CNNs trained on synthetic datasets when tested on more realistic imagery. To solve the domain shift problem, a novel augmentation technique focused on texture randomization was introduced, aimed at improving the CNN robustness against previously unseen target textures. The results prove an increase in robustness towards realistic imagery, as randomizing the texture of the target spacecraft during training allows the CNN to generalize textures and to focus on the shape of the target. However, a performance decrease in highly adverse illumination conditions or low camera exposures suggests that additional augmentation techniques are required to tackle the domain shift from an illumination standpoint.

In response to this need and in order to extend the on-ground validation to the entire navigation system, this research work proceeds by introducing the on-ground validation of a CNN-based Unscented Kalman Filter. The validation is carried out at Stanford's robotic Testbed for Rendezvous and Optical Navigation on a dataset of realistic laboratory images, which simulate rendezvous trajectories of a servicer spacecraft to the Tango spacecraft from the PRISMA mission. The validation is performed at different levels of the navigation system by first training and testing the adopted CNN on SPEED+, the next generation spacecraft pose estimation dataset with specific emphasis on domain shift between a synthetic domain and a laboratory domain. A novel data augmentation scheme based on light randomization is proposed to improve the CNN robustness under adverse viewing conditions. Next, the entire navigation system is tested on two representative rendezvous trajectories. Results indicate that the inclusion of a new scheme to adaptively scale the heatmaps-based measurement error covariance improves filter robustness by returning centimeter-level position errors and moderate attitude accuracies at steady-state. Thanks to the proposed adaptive method, the filter does not diverge in periods of low measurements accuracy, suggesting that a proper representation of the measurements uncertainty combined with an adaptive measurement error covariance is key in improving the navigation robustness.

# SAMENVATTING

Activiteiten in de ruimte zijn een nieuw tijdperk van groei ingegaan, waarbij de menselijke ontwikkeling wordt bevorderd en belangrijke aardse toepassingen zoals teledetectie, navigatie en telecommunicatie worden verbeterd. De recente ontwikkeling van SpaceX's Starlink-constellatie en de sterke toename van de lanceringen van CubeSats zullen naar verwachting een revolutie teweegbrengen in de manier waarop we de ruimte gebruiken en de huidige mogelijkheden van op satellieten gebaseerde technologie uitbreiden. Deze sterke toename van het aantal door mensen gemaakte objecten leidt echter snel tot grotere botsing risico's in overbelaste banen om de aarde. Dit heeft geleid tot de vraag of deze trend op de lange termijn houdbaar is, en uiteindelijk tot de noodzaak om duurzaamheid in de ruimte aan te pakken.

De afgelopen tien jaar hebben ruimtevaartagentschappen aanzienlijke inspanningen geleverd om zowel grote botsingen in een baan om de aarde te voorkomen via Active Debris Removal (ADR)-missies (ADR) als om de levensduur van de functionerende satellieten te verlengen met On-Orbit Servicing (OOS). Helaas wordt het naderen en vangen van ruimtepuin-objecten bemoeilijkt door het feit dat deze doelen niet meewerken en geen operaties op korte afstand kunnen ondersteunen, wat leidt tot kritieke uitdagingen bij het inschatten van hun relatieve positie en houding (pose) ten opzichte van het servicer-ruimtevaartuig . Verschillende missies zijn voorgesteld als technologiedemonstraties van puinverwijderings- en onderhoudstechnologieën, waarbij passieve monoculaire camera's worden gecombineerd met actieve sensoren om de robuustheid en nauwkeurigheid van het navigatiesysteem te verbeteren. Maar ondanks de inherente uitdagingen die gepaard gaan met het gebruik van monoculaire camera's in de ruimte, worden navigatiesystemen op basis van een enkele camera een aantrekkelijk alternatief voor systemen op basis van actieve sensoren, vanwege hun verminderde massa, stroomverbruik en systeem complexiteit. Het onderzoekswerk dat in dit proefschrift wordt gepresenteerd, is gericht op het ontwikkelen en valideren van een robuust en nauwkeurig monoculair camera-gebaseerd schattingssysteem voor relatieve poses dat voldoet aan de navigatie vereisten van zowel ADR- als OOS-missies.
Twee fundamentele open uitdagingen worden aangepakt:

1. De robuustheid en toepasbaarheid van beeldverwerking algoritmen en pose inschattingsmethoden.

2. De validatie van relatieve navigatie filters en hun interface met beeldverwerking en pose inschatting.

Dit onderzoek begint met een onderzoek naar de robuustheid en toepasbaarheid van bestaande monoculaire visie-gebaseerde relatieve pose schattingssysteem. Na het identificeren van de kenmerken en beperkingen van elk subsysteem geïmplementeerd in

state-of-the-art architecture, wordt een vergelijkende beoordeling van de huidige oplossingen gegeven op verschillende niveaus van het pose-inschattingsproces, om een nieuw en breed perspectief te bieden. Speciale aandacht wordt besteed aan de verbeterde robuustheid van nieuwe beeldverwerking schema's en pose schatters op basis van convolutionele neurale netwerken (CNN). De beperkingen en nadelen van de validatie van huidige pose-schatting schema's met synthetische afbeeldingen worden verder besproken, samen met de kritische afwegingen voor de selectie van op visuele gebaseerde navigatie filters.

Voortbouwend op de resultaten van het onderzoek wordt een nieuw raamwerk geïntroduceerd om een robuuste en nauwkeurige schatting van de relatieve pose mogelijk te maken. Twee onderzochte CNN's worden op beeld verwerkingsniveau gebruikt om een reeks vooraf geselecteerde functies op het doelruimtevaartuig te identificeren, die worden toegevoerd aan een pose-schatter voorafgaand aan het navigatie filter (los gekoppeld) of rechtstreeks aan het navigatiefilter als metingen (strak -gekoppeld). Een nieuwe methode om covariantie-matrices rechtstreeks uit de CNN-heatmaps af te leiden, wordt geïntroduceerd om de modellering van de onzekerheid van kenmerkdetectie voorafgaand aan de relatieve pose-schatting te verbeteren. De prestatieresultaten geven aan dat een nauw gekoppelde benadering een voordelige koppeling tussen de rotatie- en translatietoestanden binnen het filter kan garanderen, terwijl een representatieve covariantie van de metingen wordt weergegeven. Synthetische monoculaire beelden van het Envisat-ruimtevaartuig van de European Space Agency worden gebruikt om datasets te genereren voor training, validatie en testen van de CNN. Evenzo worden de afbeeldingen gebruikt om een representatief close-proximity-scenario na te bootsen voor de validatie van het voorgestelde filter.

Dit onderzoekswerk breidt vervolgens de validatie uit van een puur synthetische naar een meer uitgebreide validatie op de grond. Hiertoe wordt het GNC Rendezvous, Approach and Landing Simulator-testbed van de ESA gebruikt om het voorgestelde CNN-gebaseerde relatieve pose-schattingssysteem te valideren op representatieve rendez-vous scenario's, met speciale aandacht voor het oplossen van het domein verschuiving probleem dat kenmerkend is voor CNN's die getraind zijn op synthetische datasets wanneer ze worden getest op meer realistische beelden. Om het domein verschuiving probleem op te lossen, werd een nieuwe augmentatie techniek geïntroduceerd, gericht op textuur randomisatie, gericht op het verbeteren van de CNN-robuustheid tegen voorheen onzichtbare doel texturen. De resultaten bewijzen een toename in robuustheid naar realistische beelden, omdat de textuur van het doelruimtevaartuig tijdens de training allemaal willekeurig wordt verdeeld. Een prestatievermindering in zeer ongunstige verlichtingsomstandigheden of lage camerabelichting suggereert echter dat aanvullende augmentatietechnieken nodig zijn om de domeinverschuiving vanuit een verlichtingsstandpunt aan te pakken.

Als antwoord op deze behoefte en om de validatie op de grond uit te breiden tot het gehele navigatiesysteem, wordt dit onderzoekswerk voortgezet door de introductie van de validatie op de grond van een op CNN gebaseerd Unscented Kalman-filter. De validatie wordt uitgevoerd op Stanford's robot Testbed voor Rendezvous en Optical Navigation op een dataset van realistische laboratoriumbeelden, die rendez-vous-trajecten van een servicer-ruimtevaartuig naar het Tango-ruimtevaartuig van de PRISMA-missie simuleren.

De validatie wordt uitgevoerd op verschillende niveaus van het navigatiesysteem door eerst de geadopteerde CNN op SPEED+ te trainen en te testen, de volgende generatie gegevensset voor het schatten van ruimtevaartuigen met specifieke nadruk op domeinverschuiving tussen een synthetisch domein en een laboratoriumdomein. Er wordt een nieuw schema voor gegevensvergroting voorgesteld op basis van lichtrandomisatie om de CNN-robuustheid onder ongunstige kijkomstandigheden te verbeteren. Vervolgens wordt het hele navigatiesysteem getest op twee representatieve rendez-voustrajecten. De resultaten geven aan dat de opname van een nieuw schema om de op heatmaps gebaseerde meetfoutcovariantie adaptief te schalen, de robuustheid van het filter verbetert door positiefouten op centimeterniveau en gematigde houdingsnauwkeurigheden bij steady-state te retourneren. Dankzij de voorgestelde adaptieve methode divergeert het filter niet in perioden van lage meetnauwkeurigheid, wat suggereert dat een juiste weergave van de meetonzekerheid in combinatie met een adaptieve meetfoutcovariantie essentieel is voor het verbeteren van de navigatierobuustheid.

# 1

# INTRODUCTION AND MOTIVATION

*There are so many problems to solve on this planet first*
*before we begin to trash other worlds.*

E.A. Bucchianeri

Nowadays, more than 30,000 monitored pieces of debris are orbiting the Earth, including non-functional spacecraft, abandoned launch vehicle stages, and fragmentation debris[1]. Altogether, this debris poses a threat to satellites as well as human spaceflight, undermining the safety and operations in orbit. The current debris situation is a result of several factors, including a lack of mitigation strategies at the early stage of the Space Age and the absence of clear guidelines and regulations regarding post-mission disposal of inactive satellites. At the same time, the ongoing trend of increasing the number of artificial objects in space is severely increasing the risk of collisions, leading to an increased operational effort in performing collision avoidance manoeuvres. As summarized by several works (Boley and Byers, 2021; McDowell, 2020; Pardini and Anselmo, 2020), the total amount of active launches per year has increased from ~1,000 in 2011 to almost 5,000 in 2021, following a steep increase in CubeSats launches into Low Earth Orbit (LEO) and the recent creation of SpaceX's Starlink constellation (Figure 1.1).

**The Space Debris Problem** As already predicted by Kessler and Cour-Palais (1978), the increase in the number of artificial objects in Earth orbit has led to an increase in the probability of collisions between satellites. Known as *Kessler syndrome*, this phenomenon implies that even with no further launches, a single collision could rapidly trigger an avalanche effect and result in a series of inevitable collisions which will hinder access to space. Unfortunately, collisions in LEO orbits already occurred on several occasions, e.g. the Chinese ASAT test in 2007 (Thiele, 2021) and the collision of Iridium-Cosmos satellites in 2009 (Kozlowsky and Kosonocky, 2010). These events produced orbiting fragments

[1] https://www.esa.int/Safety_Security/Space_Debris/Space_debris_by_the_numbers

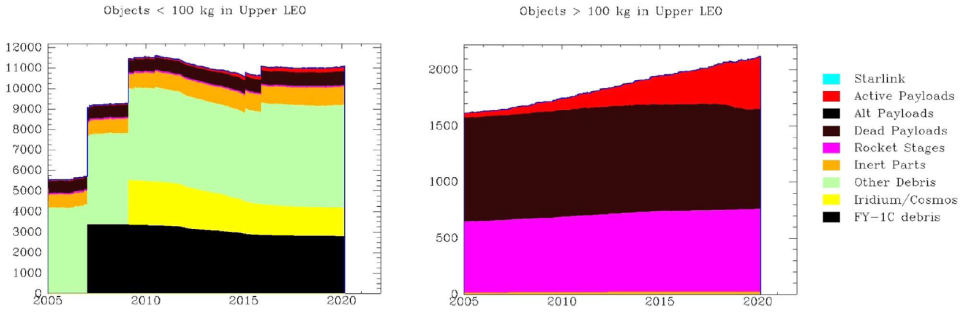



Figure 1.1: Evolution of tracked artificial Earth satellite population in the 2005–2020 period, in upper LEO (600 to 2000 km). Left: small (<100 kg) satellites. Right: large (>100 kg) satellites. The small object population is dominated by debris objects; large objects are mostly dead payloads and rocket stages. Evolution shows steady growth. Figure taken from McDowell (2020)

which increased the probability of further collisions, leading to a substantial growth of debris around the Earth. With more launches in the upcoming years leading to a further increase in the number of satellites in Earth orbits, three main strategies currently stand out as options to block this chain reaction and avoid an increase in the debris population in key orbits:

1. the introduction and enforcement of debris mitigation and post-mission disposal guidelines for future space missions,
2. the removal of the largest and most dangerous inactive satellites in Earth orbits,
3. the lifetime extension of space objects.

Although mitigation and post-mission disposal guidelines are paramount for a sustainable space in the future, Active Debris Removal (ADR) missions are considered essential to autonomously decrease the risk of collision with existing debris by removing the largest and most dangerous inactive satellites in LEO. Complementing this growing need of removing non-functional objects from the operationally important orbit regimes, autonomous refueling and repairing, known as On Orbit Servicing (OOS), is also quickly becoming the most viable solution to the debris problem by extending the lifetime of existing and planned active satellites in orbit (Flores-Abad et al., 2014; Long et al., 2007; Tatsch et al., 2006).

In recent years, several missions have been proposed as technology demonstrators of debris removal and servicing technologies. The European Space Agency's e.Deorbit mission (Wieser et al., 2015) was originally planned to remove the Envisat spacecraft from LEO, although funding of the mission stopped in 2018. As a result, ESA commissioned the ClearSpace-1 mission (Biesbroek et al., 2021), which will target the Vespa (Vega Secondary Payload Adapter) upper stage left in an approximately 800 km by 660 km altitude orbit after the second flight of ESA's Vega launcher in 2013[2]. In conjuction with ESA's efforts,

[2] https://www.esa.int/Safety_Security/Clean_Space/ESA_commissions_world_s_first_space_debris_removal

private companies have already shown ADR missions capabilities in-flight by successfully demonstrating the capturing of client spacecraft that were sent in orbit together with their main servicing satellites. In the ADR framework, Astroscale's End-of-Life Services by Astroscale-demonstration (ELSA-d) and Airbus's RemoveDEBRIS were recently designed, built and launched to successfully demonstrated a net-based capture (Aglietti et al., 2020; Forshaw et al., 2020; Forshaw et al., 2016) and a magnetic-based capture (Blackerby et al., 2019) of a client spacecraft by the main spacecraft. Additionally, SpaceLogistics' Mission Extension Vehicle spacecraft (MEV-1/-2) successfully rendezvoused and docked with Intelsat 901 and Intelsat 1002 in 2019/2020, performing the first re-positioning and life-extension services to Geostationary (GEO) satellites unprepared for docking (Pyrak and Anderson, 2022; Stadnyk et al., 2021).

## 1.1. UNDERSTANDING ADR/OOS MISSIONS

In their broadest definition, ADR/OOS missions are characterized by the removal/refueling of inactive/malfunctioning objects by an active servicer spacecraft (Figure 1.2). Similar to other rendezvous missions involving two space objects, these missions are characterized by the following main phases (Colmenarejo et al., 2013):

- **Phasing** This phase typically consists of estimating orbital parameters of the target's orbit and then aligning the orbital plane of the servicer spacecraft with the orbital plane of the target.
- **Approach** In this phase, the servicer approaches the target from typically a few kilometers down to a few meters, performing several proximity operations such as fly-around, inspections and close-approach. If the target is tumbling, the end of this phase is characterized by either a de-tumbling of the target or an angular synchronization which aligns the servicer with the target's rotational axis.
- **Capture/Refueling** This phase is characterized by a rigid or non-rigid capture of the target object by the servicer spacecraft. Depending on the main mission objective, deorbiting (ADR) or servicing (OOS) of the target occur.

While phasing generally entails *absolute* manoeuvres of the servicer spacecraft to inject it into the target's orbit with ground-in-the-loop operations, the approach and capture/refueling phases are related to the *relative* motion between the servicer spacecraft and the target object at far/close distances, typically requiring a different set of Guidance, Navigation, and Control (GNC) tasks and a special focus on *autonomy*.

**Autonomous Navigation** The need of autonomous systems in future ADR/OOS missions stems from the challenging fast dynamics involved in both the removal and servicing of orbiting objects when an active servicer spacecraft is utilized. When in close-proximity with a fast-moving target, the approach phase of the servicer is in fact an essential step before the actual joining of the two space vehicles. In this phase, the latency in ground-based operations during approach together with the sparsity of ground stations would undermine close-proximity operations. Hence, advancements in the field of autonomous

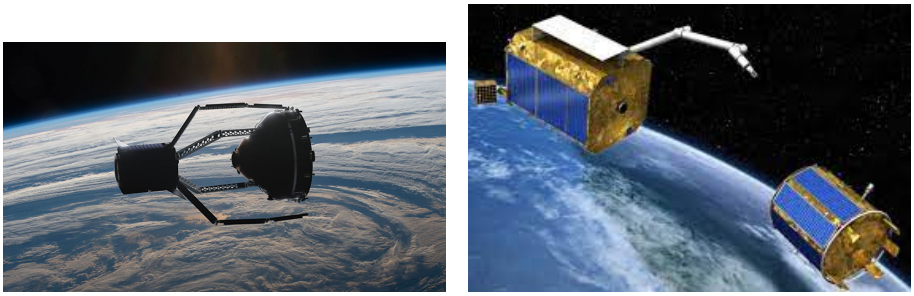
Table 1.1: Typical ADR and OOS pose requirements during approach with an uncooperative target.

| Parameter Performance | at 10 m | at 100 m (Mitchell, 2011) | at 100 m (Telaar et al., 2017) |
|---|---|---|---|
| Relative Attitude [deg] | 0.3 | 0.3 | 5 |
| Relative Rotational velocity [deg/s] | 0.1 | 0.2 | 0.5 |
| Relative Position [m] | 0.2 | 5 | 10 |
| Relative Translational Velocity [m/s] | 0.01 | 0.02 | 0.1 |

navigation are needed to extend the existing technologies and enable safe, accurate and robust methods for a successful approach. In this context, the estimation of the pose (relative position and attitude) of a target object by an active servicer spacecraft represents a critical navigation task. Notably, the target object in ADR/OOS missions is typically *uncooperative*, meaning that it does not aid the relative navigation and may not be functional. As opposed to cooperative close-proximity missions, the pose estimation problem is thus complicated by the fact that the target cannot cooperate with the servicer spacecraft during relative navigation with devices such as visual markers, Light-Emitting Diodes (LED) or corner reflectors (Opromolla et al., 2017b).

**Navigation Requirements** In terms of navigation performance requirements, there is currently no established benchmark for ADR/OOS missions. Although the phasing of the servicer spacecraft has been an essential phase in several past rendezvous missions and is well-established in the literature (Dutta and Tsiotras, 2009; Fehse, 2003), requirements for the approach phase are highly specific to the overall GNC system used during the mission, let aside the specific mission objective (i.e. capture/refueling). However, drawing from the current knowledge of planned missions, Table 1.1 shows a set of performance requirements expected from the mission concepts outlined in Mitchell (2011) in a general sense and in Telaar et al. (2017) for the Envisat target case. The requirements are reported only for the approach phase in order to generalize them to both ADR and OOS missions. Although mission- and target-specific, these requirements can be used as a reference while verifying an autonomous navigation system.

**Monocular cameras: a promising paradigm** When approaching an uncooperative target, optical sensors on the servicer spacecraft are well suited to pose estimation since they do not rely on any sensing hardware on the target, such as Global Navigation Satellite Systems (GNSS) sensors and antennas. In recent years, low Size-Mass-Power-Cost (SMaP-C) sensors became particularly suited to the limited on-board capacity of small servicing satellites, giving the rise to pose estimation systems based solely on a monocular camera. However, sensor suites in planned and flown technology demonstrator missions tend to combine monocular and stereo cameras with a Laser Detection and Ranging (LIDAR) instrument to cope with the limitations of a monocular-only sensor suite. Despite the inherent challenges that come together with the use of monocular cameras in space



(a) ClearSpace's ADR Concept (Biesbroek et al., 2021)

(b) DEOS' OOS concept (Reintsema et al., 2010)

Figure 1.2: Artistic visualization of ADR (a) and OOS (b) mission concepts involving a servicer spacecraft and a target object.

without additional sensors, the already proven benefits of these systems for Earth-based applications recently opened up to an extensive research in the field of monocular vision-based systems for future cost-effective ADR/OOS missions (Opromolla et al., 2017b). In this context, the flight results of the Autonomous Vision Approach Navigation and Target Identification (AVANTI) in-orbit demonstration mission (Gaias and Ardaens, 2018) recently explored the opportunities of a navigation system entirely based on a monocular camera, providing an in-flight validation of autonomous navigation relative to the BEESAT-4 CubeSat, from far-range down to circa 50 m of inter-satellite distance. Unfortunately, the relative navigation system was validated on relative distances for which the shape of target spacecraft could not be inferred from the monocular image, hence focusing on a mission phase where the estimation of the relative position (*angles-only* relative navigation) takes place instead of pose estimation. As a result, there is currently a need to research on, and validate, monocular-based pose estimation systems prior to assessing the in-flight capabilities of navigating with a single camera in close-proximity to an uncooperative target.

## 1.2. MONOCULAR VISION-BASED POSE ESTIMATION: A BRIEF OVERVIEW

The concept of estimation in relative navigation refers to the calculation of navigation-related quantities from direct observation in the form of measurements acquired from one or more sensors. From a high-level perspective, optical sensors suited for relative navigation can be divided into active and passive devices, depending on whether they supply their own illumination source, i.e. LIght Detection And Ranging (LIDAR) sensors and Time-Of-Flight (TOF) cameras, or if they passively acquire light, i.e. monocular and stereo cameras. Spacecraft relative navigation usually exploits Electro-Optical (EO) sensors such as stereo cameras (Davis and Pernicka, 2019; Pesce et al., 2017) and a LIDAR sensor (Opromolla et al., 2015) in combination with one or more monocular cameras, in order to overcome the partial observability that would result from the lack of range

information if a single monocular camera is used (Segal et al., 2014). In this context, monocular vision-based pose estimation refers to the estimation of the pose of a target spacecraft with respect to the servicer spacecraft by only using 2D images, either taken by a monocular camera or fused from several monocular cameras.

**Image Processing and Pose Estimation** Pose estimation systems based solely on a monocular camera are recently becoming an attractive alternative to systems based on active sensors or stereo cameras, due to their reduced mass, power consumption and system complexity (Pasqualetto Cassinis et al., 2019; Sharma, Ventura, et al., 2018). However, a significant effort is still required to comply with most of the demanding requirements for a robust and accurate monocular vision-based relative navigation system. Since the extraction of visual features is an essential step in the pose estimation process when the target is uncooperative, advanced image processing techniques are required to extract keypoints (or interest points), corners, and edges on the target body. The detected features are then matched with features on a wireframe 3D model of the target to solve for the pose. Depending on whether the 3D model of the target is available offline or if it has to be reconstructed on-board, the system can be model-based or model-free (Pasqualetto Cassinis et al., 2019). Regardless of which pose estimation method is chosen, a reliable detection of key features is critical to guarantee safe operations around an uncooperative target.

**Observation Challenges** The extraction of target features from a monocular image during image processing is a complex task which can be jeopardized by external factors, such as adverse illumination conditions in orbit, low Signal-to-Noise ratio (SNR) and Earth in the background, as well as by target-specific factors, such as the presence of complex textures and features on the target body (Pasqualetto Cassinis et al., 2022a; Pasqualetto Cassinis et al., 2019). Moreover, most of the feature extraction methods are based on the image gradient, detecting textured-rich features or highly visible parts of the target silhouette. As such, the detected features are image-specific and can vary in number and typology depending on the image histogram. This means that most techniques cannot accommodate an offline feature selection step, resulting in a computationally expensive image-to-model correspondence process to ensure that each detected 2D feature is matched with its 3D counterpart on the wireframe model of the target object. These challenges dictate that a careful selection of both image processing algorithms and pose estimation methods is paramount to guarantee a robust and reliable performance in orbit under favourable as well as highly adverse scenarios.

**Pose Initialization, Tracking and Filtering** The way the pose estimation problem is solved depends on a variety of factors, including the information available prior to the actual estimation. If *no* a-priori information of the relative state between the uncooperative target and the monocular camera is available (so called *lost-in-space scenario*), only an initial estimate of the pose can be obtained. This is referred to as *pose initialization*. Once this initial estimate is computed, *pose tracking* can be performed by collecting new camera images and using the previous estimate as a-priori information. Note that

the system performance during initialization and tracking can differ considerably. Typically, the expected pose accuracies during pose initialization are relatively low due to the challenges involved in having no a-priori knowledge of the relative state. Conversely, during pose tracking the accuracies can be refined from an initial pose estimate. As a result, more relaxed pose estimation requirements usually apply to lost-in-space scenarios, with more stringent requirements once pose initialization is achieved. Notably, both pose initialization and pose tracking are not well suited to produce pose estimates at high frequencies, especially due to the computationally expensive image processing in combination with pose estimation. Furthermore, solving for the pose solely from measurements can only provide a prediction from sensor data without accounting for any modeling of the external environment. In other words, it is not guaranteed that the estimation can reliably deal with unwanted components in the measurements. Finally, quantities such as the translational and rotational velocities can hardly be estimated together with the pose. Therefore, filtering techniques are usually used in combination with the camera measurements and the actual pose estimate in order to return full-state (pose and relative velocities) solutions at high frequency (Sharma and D'Amico, 2017). In this context, the modeling of the relative dynamics inside the filter can improve the accuracy of the predicted relative state from measurements and allow a more robust pose tracking. Essentially, the inclusion of a navigation filter results in a reduction of unwanted components in the measurements by fusing the sensors' output with the filter's internal dynamics to estimate the relative state.

In summary, a monocular vision-based pose estimation system typically consists of the following subsystems:

- **Image Processing Algorithm** which detects and extracts visual features from one or more monocular images.
- **Pose Estimator** which estimates the pose from features detected in a single image (pose initialization) or from a sequence of images (pose tracking).
- **Navigation Filter** for the estimation of the full relative state at high frequencies, i.e. including translational and rotational velocities.

Note that in this thesis a distinction is made between a pose estimation *method* (which consists of a combination of image processing algorithms and pose estimation solvers) and a pose estimation *system* (also called navigation system, which refers to the entire estimation pipeline including a navigation filter). Furthermore, it is important to mention that a clear distinction between each subsystem might not always hold for any pose estimation system. In reality, several methods can be found in which some subsystems may in fact incorporate others by performing multiple tasks.

## 1.3. RESEARCH OBJECTIVE

In this thesis, two fundamental open challenges in monocular vision-based pose estimation systems are addressed:

1. The robustness and applicability of image processing algorithms and pose estimation methods for a wide range of uncooperative targets, their dynamic states and illumination conditions.

2. The validation of relative navigation filters and their interface with image processing and pose estimation.

In this framework, the following main research question is formulated:

**Guiding Research Question**

Which combination of image processing, pose estimation and navigation filter subsystems can return a robust and accurate monocular vision-based pose estimation system compliant with navigation requirements of ADR/OOS missions?

which brings to the following Main Research Objective:

**Main Research Objective**

To develop and validate a robust and accurate monocular vision-based pose estimation system compliant with navigation requirements of ADR/OOS missions

In relation to the typical phases of ADR/OOS missions described in Section 1.1, only the approach phase with a target object will be investigated in this thesis, assuming an already-executed phasing phase as well as neglecting the challenges involved in the docking/berthing. Also, the main focus of this thesis will be on pose estimation systems which can rely on the a-priori information of the target's shape. Although this assumption might not be valid in case of debris or unknown targets, it can be expected that a close-proximity mission around the uncooperative target would accommodate a so-called *inspection* phase, in which several images of the target can be acquired and post-processed to create a representative geometric model online which can be relied upon during the pose estimation.

## 1.4. RESEARCH QUESTIONS AND THESIS OUTLINE

At the beginning of any scientific research, it is key to investigate existing solutions to the problem at hand while identifying the applicability, limitations and challenges of state-of-the-art methods. In this context, this thesis originates from the formulation of the following first research question:

**Research Question #1**

What are the characteristics and limitations of monocular vision-based systems for the pose estimation of uncooperative spacecraft?

To answer this research question, an extensive literature review is carried out in Chapter 2 at different levels of a monocular vision-based pose estimation system, in order to

identify the characteristics and limitations of each subsystem implemented in state-of-the-art architectures.

Recent surveys on the topic focused on a comparative assessment of pose estimation algorithms (Sharma and D'Amico, 2015) or provided a broader review on cooperative as well as uncooperative targets by including monocular- as well as stereo- and LIDAR-based systems (Opromolla et al., 2017b). Furthermore, only monocular cameras operating in the visible spectrum were reviewed, and recent estimation methods based on deep learning techniques were not included. The objective of the literature review performed in this work is to extend the previous surveys in mainly three aspects. Firstly, the applicability of Visible (VIS), Near-Infrared (NIR) and Thermal-Infrared (TIR) cameras is investigated. Special focus is put on multispectral data fusion, in which two or more cameras are used in different spectral ranges to improve the camera measurements. Secondly, both image processing systems and pose estimation algorithms are analyzed with particular emphasis on the relative range they were tested on, the robustness with respect to the image background, and on the synthetic (using computer renders) and real (using realistic mockups) datasets adopted for their validation. Furthermore, novel pose estimation methods are reviewed which are based on Convolutional Neural Networks (CNN), deep-learning algorithms capable of estimating the pose of a target object from a 2D image by performing a convolution operation between the input image and their network layers. Finally, a review is presented for the navigation filters currently adopted. A distinction is made between known targets, for which mass and inertia properties as well as a 3D model of the target are known and available, and partially known targets, for which the uncertainty is constrained to the target center of mass and moment of inertia, while a 3D model of the target is available. Notably, this distinction affects the internal dynamics of the navigation filter rather than on the image processing and pose estimation prior to the filter.

### 1.4.1. ROBUSTNESS AND ACCURACY: TWO KEY PERFORMANCE FACTORS

After identifying the most relevant characteristics and limitations of monocular vision-based pose estimation systems, the subsequent research aims at addressing the achievable accuracy and robustness of the image processing and pose estimation subsystems by proposing novel methods and pipelines which improve on existing systems:

**Research Question #2**

Which pose estimation methods return the robustness and accuracy required for the pose estimation of an uncooperative spacecraft?

High robustness in a pose estimation system is essential to guarantee that the system can perform under adverse orbital conditions or system failures without jeopardizing the mission objectives, e.g. by leading to a collision with the target object or to a large deviation from the desired orbit. On the other hand, high system accuracy is essential in order to guarantee the fulfillment of the strict mission requirements involved in the approach of the target object, i.e. precise relative navigation in close-proximity before

docking. Leveraging on the promising robustness and accuracy found in existing CNN-based systems, the first part of Chapter 3 focuses on the applicability of CNNs for the features detection task at image processing level, as well as on the interface of a CNN-based image processing system with existing pose estimation solvers. To validate the proposed systems, computer-generated synthetic renderings are used to simulate monocular images at representative camera-target relative ranges. Ideal as well as adverse orbital conditions (i.e. illumination conditions and camera views of the target) are simulated to stress-test the pose estimation performance in synthetic scenarios, with particular importance given to the assessment of the achievable accuracy of the proposed CNN-based system in representative ADR/OOS approach scenarios.

### 1.4.2. LEVERAGING CNNS IN VISUAL-BASED FILTERS

Once CNNs are identified as a viable solution for the task of pose estimation of uncooperative targets, their role in the overall navigation system is assessed in the second part of Chapter 3 by analyzing their interface with navigation filters:

**Research Question #3**

Which characteristics are critical in a navigation filter for the pose estimation of an uncooperative spacecraft?

Several navigation filters for close-proximity operations were investigated in recent years in the context of pose estimation (Naasz et al., 2009; Pesce, Haydar, et al., 2019; Sharma and D'Amico, 2017). However, limited attention has been given to the interface between a CNN-based system and a navigation filter. From a filter architecture standpoint, it is critical to evaluate the applicability of CNNs in so-called *tightly-coupled* and *loosely-coupled* approaches. In a loosely-coupled approach the detected features are transformed into an estimated pose prior to the filter, whereas a tightly-coupled approach directly feeds the detected target features into the filter. Moreover, it is critical to investigate the accuracies achievable by both architectures whilst stress-testing the filter robustness towards adverse conditions. To this end, the previous investigations on pose estimation are extended from the generation of static images of the target object to the generation of image sequences representative of typical close-proximity trajectories in ADR/OOS missions.

**A new method to model uncertainty in CNN-based systems** In the context of vision-based filters for pose estimation, limited research has been carried out on how to model the features detection uncertainty within the navigation filter. In existing methods, the derivation of statistical information from sensor measurements is a lengthy process which generally cannot be directly related to the actual detection uncertainty (Cui et al., 2019; Harvard et al., 2020). Moreover, standard procedures cannot be easily applied if CNNs are used in the feature detection step, due to the difficulty to associate statistical meaning to the image processing tasks performed within the network. In this context, a novel method is presented in this thesis in which the output of the CNNs is directly exploited to return relevant statistical information about the detection step. This method consists of creating

a statistical distribution around CNN detections, and to use this dispersion to derive a representative measurement error covariance matrix from each detected feature.

### 1.4.3. TOWARDS AN END-TO-END VALIDATION FRAMEWORK

After addressing the previous research questions, a CNN-based pose estimation system is proposed together with a tightly-coupled navigation filter. However, the validation carried out insofar involved computer-generated synthetic renderings of target spacecraft. Although such validation is essential to address several aspects of the system (i.e. system architecture and subsystem interface), it is evident that the system performance in space cannot be addressed when synthetic measurements are used in the validation. As already pointed out in several previous works (Kisantal et al., 2020; Schnitzer et al., 2017), the accuracy and robustness of pose estimation systems tested solely on synthetic imagery is unpredictable when the same systems are adopted in space. More specifically, the performance of standard image processing algorithms is usually optimized on a synthetic image domain, leading to a poor performance when using actual space imagery. Furthermore and aside from standard methods, CNN-based systems often need to be trained with synthetic renderings of the available target model due to a lack of availability of representative space images. As a result, their performance on more realistic images is, as well, usually unknown and difficult to predict. In other words, the synthetic datasets used to train the CNNs tend to fail in representing the textures of the target as well as the external illuminations, resulting in low pose estimation accuracies (Kisantal et al., 2020; Pasqualetto Cassinis, Menicucci, et al., 2021). Consequently, Chapters 4-5 extend the validation using synthetic images by addressing the performance of the proposed pose estimation system on realistic space imagery:

**Research Question #4**

How can the performance of monocular vision-based pose estimation systems on actual space imagery be improved?

In this context, three desirable aspects and challenges stand out for the intended validation: first of all, a proper on-ground validation framework shall be sought to test the robustness of the proposed CNN-based system against representative images of the target spacecraft, generated in a laboratory environment which simulates space-like conditions. In this thesis, this is achieved by introducing a calibration framework which returns an accurate reference for the pose between the monocular camera and the target mockup for each generated laboratory image (pose labels), in order to be able to quantify the CNN performance. Second, novel techniques shall be investigated to improve the performance of a CNN trained using synthetic images on actual space imagery. This aspect is referred to as the *domain shift problem* (Ben-David et al., 2007; Sugiyama and Muller, 2005; Tobin et al., 2017) and can be assessed at both image processing and pose estimation levels. Novel data augmentation pipelines are introduced in this thesis in which texture and light randomization effects are incorporated in the training phase of a CNN to improve its robustness towards previously unseen laboratory images of a target spacecraft. Finally, the validation shall be extended to the proposed CNN-based navigation filter by

simulating representative relative trajectories on-ground. In this context, a novel scheme to adaptively scale the CNN-based measurement error covariance based on the filter innovations is proposed and tested on the selected rendezvous trajectories to improve the navigation performance on realistic imagery.

Besides the above mentioned aspects, the validation of the proposed CNN-based pose estimation system in space is further extended by testing its implementability on a representative space processor. This is carried out by comparing the performance of the proposed CNN model on a Graphics Processing Unit (GPU) with the performance of a converted model in a processor with low-power consumption.

### 1.4.4. RESEARCH METHODOLOGY

From a high-level perspective, the research methodology followed in this thesis consists of a step-by-step validation at different levels of the pose estimation system. Existing as well as novel image processing, pose estimation, and navigation methods are implemented in representative ADR/OOS scenarios in order to investigate their applicability and select a definite pose estimation system. At an image processing level, existing CNN architectures are re-adapted from the task of human pose estimation to the task of spacecraft pose estimation. A comparative assessment is carried out between two different CNNs to evaluate their applicability to the investigated ADR scenarios. At a pose estimation level, existing pose estimators are implemented to assess the performance of a CNN-based pose estimation method. Special focus is put on the performance of the selected pose estimation method on adverse orbital scenarios. Finally, existing filters are implemented at a navigation level and re-adapted to interface with the selected CNN-based pose estimation methods. Special focus is given to the introduction of a CNN-based representation of measurements uncertainty as well as to adaptive techniques to improve the navigation robustness against adverse orbital conditions.

**Selecting a target object** This thesis focuses on the approach phase of an ADR mission around ESA's Envisat spacecraft. Envisat is currently identified as the largest and most dangerous orbital debris in LEO, thus representing the ideal target object of an ADR demonstration mission (Wieser et al., 2015). Studies agree that the Envisat spacecraft is spinning with a main rotational rate of around 2.5-3.5 degrees per second (Sagnières and Sharf, 2019). As such, the selected target represents a worst-case scenario from a navigation perspective due to the rapidly changing views of the target and the resulting challenges in tracking. Note that the validation of the proposed pose estimation system on the selected ADR scenario can be easily extended to less severe OOS scenarios in which the target object is rotating at a much slower rate.

**Benchmarking opportunities** Thanks to an increasing interest in monocular vision-based pose estimation, considerable efforts were made in the past years to create a large dataset of monocular images of a representative target space object, on which pose estimation systems could be benchmarked. This led to Stanford's Spacecraft PosE Estimation Dataset (SPEED) and its extension SPEED+ (Kisantal et al., 2020; Park, Martens, et al., 2021), the first publicly available datasets on pose estimation. The SPEED dataset

consists of 15,300 images of the Tango spacecraft from the PRISMA mission (D'Amico et al., 2013). Specifically, SPEED comprises 15,000 synthetic images and 300 simulated images captured from the robotic Testbed for Rendezvous and Optical Navigation (TRON) at Stanford's Space Rendezvous Laboratory (SLAB). Extending the full orientation space and distances of SPEED with realistic re-creation of Earth albedo and direct sunlight present in spaceborne imagery, SPEED+ includes 60,000 synthetic images and 9,531 lab-generated images, providing unique and unprecedented quantity and quality of mockup spacecraft images. Remarkably, the capabilities of the TRON facility were recently extended towards the generation of realistic imagery of two close-proximity trajectories around the Tango spacecraft with realistic illumination conditions of a typical LEO orbit, allowing the benchmark at a navigation filter level (Park and D'Amico, 2022a). Additionally, pose estimation systems can be evaluated on the PRISMA25 dataset (D'Amico et al., 2013), which consists of 25 flight images of the Tango spacecraft acquired during the the PRISMA mission. Despite the limited number of images in PRISMA25, the comparative study between SPEED+ and actual flight images allows to assess the applicability of lab-generated images as a surrogate of flight images for validation.

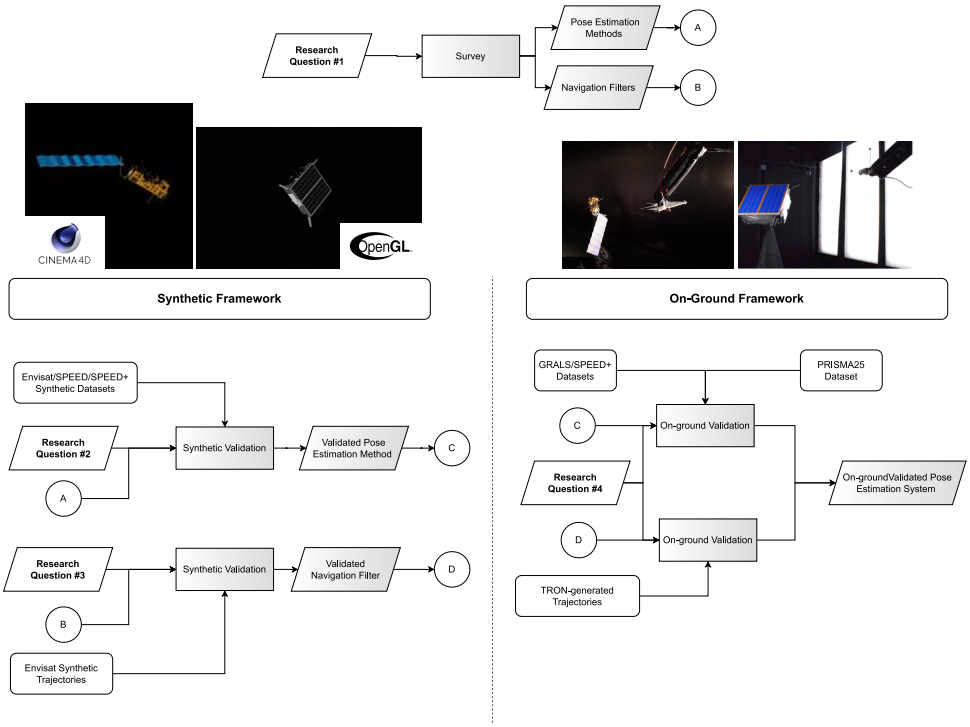



Figure 1.3: Validation Framework followed in this thesis. The validation is carried out in a sequential way at different stages of the pose estimation system, in order to stress-test each subsystem individually before validating its interfaces with other key subsystems. A laboratory framework is defined beside a synthetic one in order to assess the performance of the pose estimation system on lab-generated images as a surrogate of flight images.

Table 1.2: Research methodology-to-Research Question Validation Matrix. Here, IP, PE and NAV stand for Image Processing, Pose Estimation and Navigation, respectively.

| Research Question | System | Method | Type of Analysis |
|---|---|---|---|
| **1**: Characteristics and limitations of current pose estimation systems | IP/PE/NAV | Survey | Data Collection |
| **2**: Robustness and accuracy of pose estimation systems | IP/PE | Synthetic Validation | Simulations with Envisat imagery<br>Benchmark on SPEED synthetic images |
| **3**: Critical characteristics of navigation filters | NAV | Synthetic Validation | Simulations on Envisat trajectories<br>Benchmark on Prisma trajectories |
| **4**: System performance on actual space imagery | IP/PE<br>NAV | On-ground Validation | Analysis on SPEED+/GRALS datasets<br>Analysis on TRON-generated trajectories |

**Validation Framework** Figure 1.3 and Table 1.2 illustrate and describe the validation framework followed throughout this thesis. As already anticipated, the validation is divided into synthetic (using computer-generated renderings) and on-ground (using realistic images generated on-ground in laboratory environments) and is performed at each level of the pose estimation system. In the synthetic framework, the principal validation is carried out with images of the Envisat spacecraft rendered in the Cinema 4D© software. The adopted rendering model includes several key material properties and allows to simulate realistic reflections of the target surfaces as well as other important material effects. Besides, benchmarking is performed with synthetic images of both SPEED/SPEED+ datasets of PRISMA's Tango spacecraft, rendered with OpenGL©-based Optical Stimulator (OS) camera emulator software of the Stanford's SLAB multi-Satellite Software Simulator (Park, Martens, et al., 2021). To extend the validation using synthetic images, an on-ground validation is executed in a laboratory framework to assess the applicability of the pose estimation system in space-like conditions. In the Envisat scenario, the validation makes use of the GNC Rendezvous, Approach and Landing Simulator (GRALS) testbed of the Orbital Robotics & GNC laboratory (ORGL) facility at ESA's Research and Technology Centre (ESTEC). Only a validation of the proposed pose estimation method is performed at this stage. Conversely, the benchmarking on-ground validation is performed at both pose estimation and navigation level with the realistic subset of SPEED+ as well as with realistic close-proximity trajectories recreated at Stanford's TRON facility.

### 1.4.5. THESIS CONTRIBUTIONS

The primary contribution of this research is the development and validation of a CNN-based, monocular-based system for the pose estimation of uncooperative spacecraft. This contribution is articulated into four main aspects of the proposed system:

1. A *systematic review contribution* through the review of the robustness and applicability of existing monocular vision-based pose estimation systems.

2. An *algorithmic contribution* through the introduction of a CNN-based represen-

tation of feature detection uncertainty and of an adaptive scheme to improve the statistical knowledge of the measurements within the navigation filter.

3. A *systems engineering contribution* through a detailed analysis of the interfaces between image processing, pose estimation, and navigation in the proposed monocular-based system.

4. An *experimental contribution* through the creation of a calibration framework for ESTEC's GRALS testbed and through the on-ground validation of the proposed CNN-based pose estimation system on representative lab imagery generated both at ESTEC's GRALS facility and at Stanford's TRON facility.

Note that this thesis focuses on an open-loop validation, meaning that the interfaces between the proposed navigation system and guidance and control are not accounted for at a GNC level in closed-loop. Furthermore, both synthetic and on-ground validations are complemented with analyses and simulations performed in Python© and Matlab©/Simulink©.

# 2

# MONOCULAR POSE ESTIMATION SYSTEMS: A SURVEY

*The real voyage of discovery consists not in seeking new landscapes, but in having new eyes.*

Marcel Proust

## 2.1. INTRODUCING THE POSE ESTIMATION FRAMEWORK

From a high-level perspective, a monocular pose estimation system receives as input a 2D image and matches it with an existing wireframe 3D model of the target spacecraft to estimate the pose of such target with respect to the servicer camera. Referring to Figure 2.1, the pose estimation problem consists in determining the position of the target's centre of mass $\mathbf{t}^C$ and its orientation with respect to the camera frame $C$, represented by the rotation matrix $\mathbf{R}_B^C$. The Perspective-n-Points (P$n$P) equations (Fischler and Bolles, 1980),

$$\mathbf{r}^C = \begin{bmatrix} x^C & y^C & z^C \end{bmatrix}^T = \mathbf{R}_B^C \mathbf{r}^B + \mathbf{t}^C \tag{2.1}$$

$$\mathbf{p} = [u_i, v_i] = \left[ \frac{x^C}{z^C} f_x + C_x, \frac{y^C}{z^C} f_y + C_y \right], \tag{2.2}$$

relate the unknown pose with a feature point $\mathbf{p}$ in the image plane via the relative position $\mathbf{r}^C$ of the feature with respect to the camera frame. Here, $\mathbf{r}^B$ is the point location in the 3D model, expressed in the body-frame coordinate system $B$, whereas $f_x$ and $f_y$ denote the focal lengths of the camera and $(C_x, C_y)$ is the principal point of the image, defined as the point where the line from the camera centre perpendicular to the image plane meets

Parts of this chapter have been published in Pasqualetto Cassinis et al. (2019).

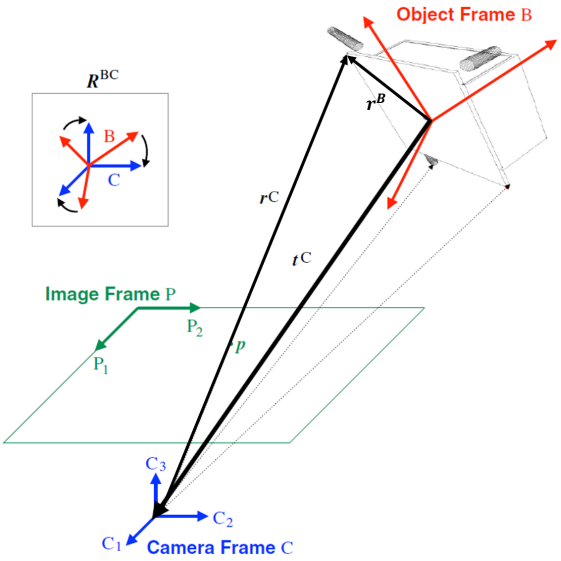



Figure 2.1: Schematic representation of the pose estimation problem using a monocular image. Figure adapted from Sharma, Ventura, et al. (2018).

the image plane. Eqs. 2.1-2.2 can be rewritten in compact form by using homogeneous coordinates,

$$\omega_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} & K & \end{bmatrix} \begin{bmatrix} & P & \end{bmatrix} \begin{bmatrix} x_i^B \\ y_i^B \\ z_i^B \end{bmatrix}, \tag{2.3}$$

where $\boldsymbol{K}$ is the intrinsic matrix of the camera and $\boldsymbol{P} = [\boldsymbol{R}_B^C | \boldsymbol{t}^C]$ is the unknown camera pose matrix. From these equations, it can already be seen that an important aspect of pose estimation resides in the capability of the image processing system to extract features $\boldsymbol{p}$ from a 2D image of the target spacecraft, which in turn need to be matched with features in the wireframe 3D model. Notably, a wireframe model of the target could either be made available offline or be reconstructed online, either prior to or during the actual estimation. Moreover, it can also be expected that the time variation of the pose plays a crucial role while navigating around the target spacecraft, e.g. if rotational synchronization with the target spacecraft is required in the final approach phase. As such, it is clear that the estimation of both the relative translational and angular velocities represent an essential step within the navigation system.

This Chapter reviews existing monocular vision-based pose estimation systems, from the adopted monocular sensors (Section 2.2) to the currently adopted image processing algorithms (Section 2.3.1), pose estimation methods (Section 2.3.2-2.3.5) and navigation architectures (Section 2.4). The aim of this survey is to provide a detailed overview of the robustness and applicability of state-of-the-art monocular-based pose estimation

systems for the relative navigation with an uncooperative target. To this end, existing surveys are extended in mainly three directions. Firstly, focus is put on the applicability and robustness of multispectral monocular cameras as opposed to sensor suites in which only monocular cameras operating in the visible spectrum are used. Secondly, both image processing and pose estimation algorithms are analyzed with particular emphasis on the relative range they were tested on, the robustness with respect to the image background, and on the synthetic and real images database adopted for their validation. Furthermore, novel pose estimation schemes are reviewed which are based on Convolutional Neural Networks (CNN), following a recent rapid increase of estimation methods based on deep learning techniques. Finally, a review is presented for the navigation filters currently adopted. A distinction is made between *known* targets, for which mass and inertia properties as well as a 3D model of the target are known and available, and *partially known* targets, for which the uncertainty is constrained to the target center of mass and moment of inertia, while a 3D model of the target is available. Note that *unknown* targets, for which the 3D model is also unknown, are not reviewed.

## 2.2. REVIEW OF MONOCULAR EO SENSORS

One of the first applications of Visible (VIS) cameras for the pose estimation of an uncooperative target is represented by the Relative Navigation Sensor which flew as part of the Hubble Space Telescope (HST) Servicing Mission 4 (SM4). The camera suite consisted of three monocular cameras operating at long, medium and short range of the visible region (Naasz et al., 2009) to aid the estimation of pose of the target telescope, assumed to be unknown. Subsequently, inspired by the promising applications of existing visual-based systems for present and future formation flying missions and OOS missions, many authors continued with the investigation of the feasibility of VIS cameras for the pose estimation of uncooperative spacecraft. Du et al. (2011) proposed a scheme which combines a singular VIS camera, in the far- and mid-range mission phases, with two collaborative monocular VIS cameras in the final approach phase, in order to increase the camera FoV and aid the feature extraction within the image processing system. The cameras were used to estimate the pose of large non-cooperative satellites in Geostationary Earth Orbit (GEO). Liu and Hu, 2014 evaluated the performance of a pose estimation method for cylinder-shaped spacecraft which makes use of single images from a monocular VIS camera, whereas other authors used images collected by the monocular VIS camera onboard the PRISMA mission to investigate the robustness of several pose estimation schemes with respect to image noise, illumination conditions and Earth in the background geometries (D'Amico et al., 2014; Sharma and D'Amico, 2017; Sharma, Ventura, et al., 2018). Furthermore, Schnitzer et al. (2017) included two monocular VIS cameras in the sensors suite adopted in their on-ground testing of image-based non-cooperative rendezvous navigation, and Pesce, Opromolla, et al. (2019) adopted a single passive monocular camera to reconstruct the pose of an uncooperative, known target. Despite the differences in the experimental setup, as well as in the pose estimation schemes, a common feature that was found for VIS cameras, even for cooperative pose estimation, is their strong dependency on the Sun's or Earth's illumination, which becomes more severe when the target does not have any fiducial marker.

Opposed to VIS cameras, Thermal-Infrared (TIR) cameras are infrared cameras sensitive to the mid- and far-infrared spectral ranges (3 $\mu$m - 14 $\mu$m). Due to size, complexity, and power consumption of cryogenically-cooled infrared sensors, the current state-of-the-art on TIR cameras for spacecraft relative navigation relies on uncooled microbolometers operating in the range 8 $\mu$m - 14 $\mu$m, as they can provide sufficient sensitivity at low cost (Kozlowsky and Kosonocky, 1995). This type of sensor was flight-tested as part of the LIRIS demonstrator during the ATV5 Mission (Cavrois et al., 2015) as well as part of the Raven ISS Hosted Payload (Galante et al., 2016), and it has been used in Yilmaz et al. (2017) as well as in Gansmann et al. (2017) and in Schnitzer et al. (2017) to assess the robustness of a TIR-based navigation system for ADR and to validate a pose estimation method based on feature extraction, respectively. Also, synthetic and real TIR camera images were recently used to validate a model-based and an appearance-based pose estimation methods (Shi et al., 2016, 2017; Shi, Ulrich, Ruel, and Anctil, 2015). Notably, the data from the TIR camera in Galante et al. (2016) were fused with the data of a visual camera and a flash LIDAR in order to improve the overall sensors performance. In a recent effort, Shi and Ulrich (2021) investigated the robustness of a TIR-based pose estimation system on synthetically generated images of the Envisat spacecraft as well as on real rendezvous flight images of the ISS. Their method included a novel foreground extraction method aimed at reducing the image processing speed that affects more traditional traditional techniques. Furthermore, Colombo et al. (2022) proposed an approach in which synthetic VIS and TIR images of the Vespa upper stage are fused at image level to obtain complementary and more informative results that can improve pose estimation in challenging illumination scenarios.

When compared to VIS cameras, TIR cameras do not depend on external light sources but rather on the emitted thermal radiation of the target spacecraft itself, thus avoiding any saturation due to Sun presence in the camera FoV or Earth in the background. This makes the sensor more robust against the different illumination conditions, typical of an ADR scenario (Deloo and Mooij, 2015). On the other hand, their image resolution is usually much lower than VIS camera. As reported in Yilmaz et al. (2017), the amount of blur in the image can significantly affect the performance of feature detection algorithms within the image processing system. Also, the results of the tests with real TIR camera images in Schnitzer et al. (2017), in which a scaled model of the Envisat was heated through resistors mounted on the rear of the plates and a Halogen lamp was used for the illumination, demonstrated that real TIR images clearly differ from synthetic images. More in particular, Barrel distorsion was found to be more severe than the one modelled in the synthetic dataset, and the edges of the spacecraft silhouette were found more faded in the real images compared to the synthetic ones. Furthermore, the different thermal dynamics encountered during an ADR mission due to varying temperature profile of the target over one orbit, as well as the different thermal surface coatings of the target, introduce some challenges in the imaging. As an example, the performance of the method proposed in Shi, Ulrich, Ruel, and Anctil (2015) cannot be evaluated due to the too optimistic assumptions of the thermal environment of the target. Furthermore, as stated in Shi et al. (2017), the resolution of TIR images sensibly affects the accuracy of the pose determination in the training phase of a model-free method.

Table 2.1: Comparison between different camera suites for space applications

| | **Saturation due to the Sun** | **Robustness w.r.t. Eclipse** | **Robustness w.r.t. Earth in background** | **Image quality** | **Robustness w.r.t thermal dynamics** |
|---|---|---|---|---|---|
| VIS | Inferior | Inferior | Inferior | Superior | Superior |
| TIR | Superior | Superior | Superior | Inferior | Inferior |
| NIR | Inferior | Superior | Inferior | Superior | Inferior |

Finally, Near-Infrared (NIR) cameras are cameras which operate in the spectral range from 780 to 2500 nm. As such, current CMOS/CCD technologies can be adopted to sense the incoming NIR radiation, and a superior image quality compared to TIR microbolometers can be achieved. To the best of the authors' knowledge, the only pose estimation scheme so far tested with NIR images is based on a model-based image processing in which the camera suite combines VIS/NIR/TIR images to increase the robustness of the pose estimation . This work was part of an ESA's Technology Research Programme (TRP) study called Multi-spectral Sensing for Relative Navigation (MSRN)[1], which focused on the design of a multispectral camera that can be used for navigation purposes in a wide variety of scenarios. This activity focused on increasing the accuracy and robustness of normal multispectral cameras by combining a Visual-Near Infra-Red (VNIR) spectral channel to a TIR spectral channel (Wieser et al., 2015). In this way, the benefits of each single camera type, listed in Table 2.1, can be combined to return a superior performance of the camera suite. Figure 2.2 illustrates the different coupling schemes proposed. Data fusion both at image and image processing levels was investigated in order to comply with the requirements of a robust and computationally fast image processing prior to the navigation filter.

Very recently, In-Orbit Demonstration (IOD) missions such as ELSA-d (Blackerby et al., 2019) and RemoveDebris (Forshaw et al., 2016, as well as the MEV-1 and MEV-2 missions (Pyrak and Anderson, 2022) which refuelled two Intelsat GEO satellites, included VIS/TIR/NIR monocular cameras in their sensor suite. However, the sensor suite also included active sensors such as radars and LIDARs, meaning that a fully monocular system was not implemented during adverse illumination conditions, Earth in the background, and near-eclipse scenarios.

The current state-of-the art on monocular cameras is further reviewed by focusing on the applicability of the proposed camera suites for the desired operational range, considering the requirement to have a robust pose estimation of an uncooperative target from several hundreds of meters down to docking, which characterises most of the close-proximity rendezvous missions. Table 2.2 lists some relevant characteristics of the camera suites and reports the tested range of the pose estimation simulations. Naasz et al. (2009)

[1] https://www.esa.int/Our_Activities/Space_Engineering_Technology/Shaping_the_Future/Multispectral_Sensing_for_Relative_Navigation

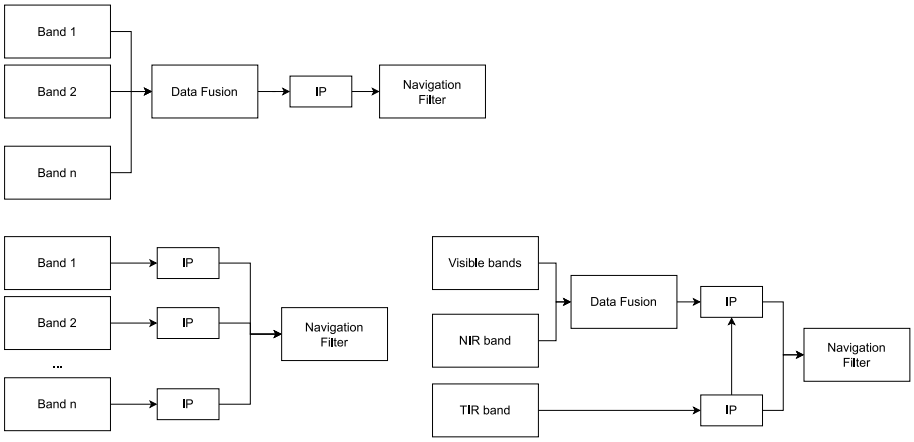



Figure 2.2: Illustration of the MSRN cameras coupling schemes[1]. Here, IP stands for Image Processing.

and Cavrois et al. (2015) tested monocular cameras down to 0.5 meters from the target and down to actual docking, respectively. However, the challenges of feature extraction within the image processing at close range were not investigated. As an example, with a FoV of around 23 degrees and a distance from the target of around 0.5 meters, the image processing would need to extract features from a portion of the spacecraft as small as a 0.2 $x$ 0.2 m rectangle, which can be challenging if the satellite is relatively large. On the other hand, the claim in Du et al. (2011) that collaborative cameras are strictly required for the close approach phase relates to the fact that their selected image processing scheme is based on the extraction of large rectangular features of large communication GEO satellites. Other authors investigated several different pose estimation schemes which rely on more flexible feature extractions. However, their pose estimation systems were not tested for relative ranges below 5 meters. It can be concluded that some effort is still required to assess whether a single monocular camera can be used for close-proximity pose estimation of an uncooperative target or if collaborative cameras are needed. As a general remark, it should in principle be possible to rely on a single monocular camera when the target is fully in the camera FoV, and switch to the feature tracking of the desired docking port for closer ranges, as performed in Schnitzer et al. (2017). Furthermore, several orbit scenarios should be simulated in future tests in order to investigate the robustness and applicability of each type of monocular camera as well as a combined VNIR/TIR camera suite for multispectral imaging. The scheme in Figure 2.2, as well as the one proposed in Galante et al. (2016) provided that no LIDAR systems are considered, shall be investigated. Finally, the infrared characteristics of the target spacecraft should be fully understood in order to maximize the performance of the NIR/TIR cameras. Although Yilmaz, Aouf, Checa, et al. (2017) proposed an infrared signature estimation method capable of characterizing the dynamical thermal behaviour of space debris, some effort is still required to assess its validity and to confirm whether an exact infrared appearance model of the target is needed for a robust relative navigation solution which relies on IR images.

Table 2.2: Characteristics of the camera suites adopted in different pose estimation schemes and their tested range.

| Ref. | Camera Suite | Tested range [m] | FoV [deg] |
|---|---|---|---|
| Naasz et al., 2009 | 3 monocular VIS cameras | 150 m - 1 m | 11/23/23 |
| Du et al., 2011 | monocular + collaborative VIS cameras | 300 m - 1 m | 55 |
| Liu and Hu, 2014 | Monocular VIS camera | 40 m - 5 m | - |
| D'Amico et al., 2014<br>Sharma and D'Amico, 2017<br>Sharma, Ventura, et al., 2018<br>Sharma, Beierle, et al., 2018 | Monocular VIS camera | 13 m - 8 m | 22.3 - 16.8 |
| Cavrois et al., 2015 | 3 Monocular VIS/TIR cameras | 70 km - 8 km<br>3.5 km - docking | 60x45 |
| Shi, Ulrich, Ruel, and Anctil, 2015,<br>Shi et al., 2016,<br>Shi et al., 2017 | Monocular TIR camera | ~5 m | 40 |
| - | Monocular VNIR/TIR camera[1] | far range - 7 m | 40x40 VNIR<br>40x30 TIR |
| Yilmaz et al., 2017 | Monocular TIR camera | - | 30 |
| Schnitzer et al., 2017 | 2 Monocular VIS/TIR cameras | 100 m - docking | - |
| Gansmann et al., 2017 | Monocular TIR camera | 70 m - 21 m | - |
| Pesce, Opromolla, et al., 2019 | Monocular VIS camera | < 30 m | - |
| Shi and Ulrich, 2021 | Monocular TIR camera | 80 m | - |
| Colombo et al., 2022 | Monocular VIS/TIR camera | <30 m | - |

## 2.3. MONOCULAR POSE ESTIMATION METHODS

A monocular pose estimation method is associated to the estimation of the pose from an input image prior to the navigation filter. From a high level perspective, the architecture of a standard pose estimation process involves an initialization step, in which there is no a-priori information on the target pose, and a tracking step, in which knowledge from the previous estimates is used when new images of the target are acquired. In both cases, estimation methods can be divided into model-based and model-free. Model-based pose estimation makes use of a simplified wireframe 3D model of the target. On the other hand, model-free methods estimate the spacecraft pose without using an existing 3D model of the target. The following methods are considered in an effort to provide a comprehensive review:

- Feature-based methods
- Appearance-based methods
- CNN-based methods.

In feature-based methods, the features extracted by image processing algorithms are fed together with a wireframe model to pose estimation solvers, in order to estimate the pose. In appearance-based methods, the pose estimation is instead performed by comparing the input image with a pre-stored database of images without using standard

Table 2.3: Characteristics of state-of-the art Image Processing algorithms. Here, NA refers to the fact that no robustness tests could be found in the reference.

| Ref. | Image Processing | Tested Range [m] | Robust w.r.t. background |
|---|---|---|---|
| Naasz et al., 2009 | Digital corr./ Sobel | 150 - 1 | NA |
| Du et al., 2011 | Canny + HT | 300 - 1 | NA |
| Liu and Hu, 2014 | Ellipses extraction | 40 - 5 | NA |
| D'Amico et al., 2014 | LPF + Canny + HT | 13 - 8 | No |
| Shi, Ulrich, Ruel, and Anctil, 2015 | RCM + HCD | ~5 | NA |
| Galante et al., 2016 | Sobel | NA | NA |
| Shi et al., 2016 | CLAHE + SIFT/ BRIEF + RANSAC | - | NA |
| Gansmann et al., 2017 | Canny | 100 - 21 | NA |
| Rondao and Aouf, 2018 | FREAK + EDL | NA | NA |
| Sharma, Ventura, et al., 2018 | WGE + S/HT | 13 - 8 | Yes |
| Pesce, Opromolla, et al., 2019 | GFTT | < 30 | NA |
| Capuano et al., 2019 | Prewitt + gradient filter ST+HT+LSD | 45 - 5 | Yes |
| Shi and Ulrich, 2021 | LAPLACE foreground mask enhanced graph manifold ranking | 80 | Yes |

image processing algorithms. As such, no feature extraction is required. Finally, in CNN-based pose estimation methods, a CNN is trained offline and exploited to estimate the pose. Depending on the adopted architecture, the CNN can be either used in place of standard image processing algorithms (feature-based), or replace the entire pose estimation pipeline to directly return an estimate of the pose. For this reason, CNN-based system are treated separately from standard feature-based methods.

### 2.3.1. IMAGE PROCESSING ALGORITHMS

Image processing is a fundamental step in feature-based pose estimation, and several methods exist in literature to extract and detect target features from a monocular 2D image, based on the specific application. From a high-level perspective, the target features can be divided into keypoints (or interest points), corners, edges and depth maps. Table 2.3 provides a list of the image processing schemes reviewed in this Section. Naasz et al. (2009) accomodated two different image processing algorithms within their RNS system: a Sobel edge-enhancing image filter to process a 10-bit camera image and perform the edge extraction, also adopted in Galante et al. (2016), and a digital correlation image processing technique which computed the position of certain features of the target spacecraft. These two methods were used separately by different pose estimation systems which were tested during the HST-SM4. Several realistic lighting conditions were simulated to validate the robustness of the image processing algorithms with respect to illumination. Du et al. (2011) included a median filter before the other steps of the image processing to cope with image noise and smooth the data. The Canny edge detection algorithm was selected

to detect edges in the image, and a subsequent Hough transform (HT) (Duda and Hart, 1972) was used to extract the detected lines. Several tests were conducted to assess the robustness of image processing with respect to image noise at different variance levels. However, a limitation of their method was that it focused on the extraction of rectangular structures on a large target spacecraft. Liu and Hu (2014) presented a robust method based on ellipses extraction for cylinder-shaped spacecraft, but its application is not feasible for the pose estimation of a spacecraft of generic shape. D'Amico et al. (2014) used the same feature detection and extraction methods described in Du et al. (2011) in combination with a Low-Pass Filter (LPF). This method was tested with the PRISMA image dataset and proved to be flexible with respect to the spacecraft shape, but it lacked of robustness to illumination and background conditions. Furthermore, it did not prove to be robust for symmetric spacecraft. Shi, Ulrich, Ruel, and Anctil (2015) selected the Roberts Cross Method (RCM) in combination with the Harris Corner Detection (HCD) method to improve the computational time of the image processing. However, the limitations of the RCM in producing less edges than the Canny's were not assessed. Shi et al. (2016) implemented a Contrast Limited Adaptive Histogram Equalization (CLAHE) to clean and restore blurred TIR images. A Scale Invariant Feature Transform (SIFT) (Lowe, 2004), in combination with the Binary Robust Independent Elementary Features (BRIEF) method (Calonder et al., 2010), was used to extract the target interest points from the denoised image. The RANdom SAmple Consensus (RANSAC) (Fischer and Bolles, 1981) algorithm was further included to rapidly extract image features and descriptors by using some internally pre-stored test image features for feature matching. Yilmaz, Aouf, Majewski, et al. (2017) performed an evaluation of the invariance of edge and corner detectors applied to TIR images. The Good Feature to Track (GFTT), Speeded Up Robust Features (SURF) and Phase Congruency Point (PC-P) edge algorithms, as well as edge detectors such as Sobel, were traded-off based on their robustness under different thermal conditions representative of the dynamic space thermal environment. Their results showed that thermal variations can cause a significant variation in the thermal signatures, and thus challenge the robustness of pose estimation methods based on feature extraction. Gansmann et al. (2017) adopted the Canny algorithm to extract edges from TIR images and from a 2D rendered representation of the target, obtained by projecting a 3D model. The variation in brightness and the variation in depth were used to extract the edges from the TIR images and from the render, respectively. Furthermore, Rondao and Aouf (2018) adopted a Fast Retina Keypoint (FREAK) descriptor in combination with the Edge Drawing Lines (EDL) detector to extract keypoints, corners, and edges to find the correspondence between features. In their method, a depth mapping was further performed which aided the feature extraction. The limitation of these two latter methods is that they require an offline database for image matching. More recently, Sharma, Ventura, et al. (2018) proposed a novel technique to eliminate the background of images, called Weak Gradient Elimination (WGE). After using a Gauss filter to blur the original image and aid the feature extraction, the image gradient intensities were computed, and the WGE was used to threshold the weak gradient intensities corresponding to the Earth in the background. In the next step, the Sobel algorithm and the Hough Transform (S/HT) were used to extract and detect features. Notably, the WGE technique can also be used to identify a rectangular region of interest (ROI) in the image which can allow an automated

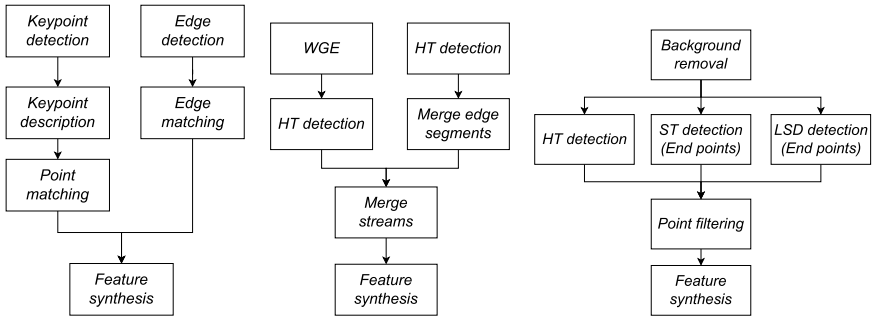



Figure 2.3: Examples of feature synthesis schemes. (a) Rondao and Aouf, 2018, (b) Sharma, Ventura, et al., 2018, (c) Capuano et al., 2019.

selection of the hyperparameters required by the HT. In this way, the hyperparameters are automatically scaled based on the varying distance from the target. By creating two parallel processing flows, the method proved to be able to extract main body features as well as particular structures such as antennas, and thus to solve the symmetry ambiguity which characterized other image processing schemes. Furthermore, the implementation of the WGE method returned a much higher robustness with respect to Earth in the background compared to the other methods. However, scenarios in which the Earth horizon is present in the background represented a challenge for the image processing due to an improper ROI detection. Recently, Capuano et al. (2019) introduced a new scheme in which three different parallel processing streams, which use the Shi-Tommasi (ST) corners detector, the HT, and the Line Segment Detector (LSD), are exploited in order to filter three sets of points and improve the robustness of the feature detection. This was performed in order to overcome the different drawbacks of each single method. Feature fusion was then used to synthesise the detected points into polylines which resemble parts of the spacecraft body. By including a background removal step similar to the WGE in Sharma, Ventura, et al. (2018), which makes use of a Prewitt operator in combination with a gradient filter, the authors could also demonstrate the robustness of their image processing with respect to the Earth in the background. Furthermore, the scenarios with the Earth horizon were tackled by tuning the threshold of gradient filter to a more selective value. The last three feature extraction schemes (Capuano et al., 2019; Rondao and Aouf, 2018; Sharma, Ventura, et al., 2018), which combine several keypoints, edges and corners detectors, are depicted in Figure 2.3. Finally, Shi and Ulrich (2021) proposed an enhanced gradient descent method to extract edges from TIR images of uncooperative targets. Their method is combined with LAPLACE, a novel foreground mask algorithm capable of efficiently separating the spacecraft from the background. The validation of the entire image processing pipeline was performed on both synthetic and real TIR images, demonstrating the robustness of image processing against challenging image backgrounds, such as Earth in the background during Nadir pointing of the servicer spacecraft.

As a general remark, image processing algorithms based on keypoint features detec-

Table 2.4: Comparative assessment results from simulations as a qualitative decision matrix in Sharma and D'Amico (2015). Here, PosIt+ refers to a solver that can switch between Coplanar PosIt and PosIt.

| Solver | Number of Features | Noise | Outliers | Distance to Camera |
|---|---|---|---|---|
| PosIt | Nominal | Superior | Inferior | Nominal |
| EP$n$P | Superior | Par | Inferior | Inferior |
| PosIt+ | Nominal | Superior | Inferior | Nominal |
| NRM | Superior | Superior | Nominal | Nominal |



tors (i.e. SURF, ORB, SIFT, FREAK) present some advantages compared to algorithms based on edge and corner detectors (i.e. HT, EDL, HCD), given their invariance to perspective, scale and illumination changes (Bay et al., 2008; Lowe, 2004). However, these features are still sensitive to extreme illumination scenarios. Moreover, their robustness with respect to outliers, which would be present i.e. when the Earth is in the image background, has not been fully proved yet in the framework of pose estimation in space. On the other hand, the recent advancements in the image processing algorithms based on corners/edges detection showed an improvement in the robustness of such methods with respect to the Earth in the background (Sharma, Ventura, et al., 2018). Furthermore, edges and corners detectors are retained to be more robust than features detectors in case of partial occlusion of the target, especially during tracking (V. Lepetit and Fua, 2005). Future works should focus on the assessment of the robustness of keypoint features detectors to outliers in space imagery, as well as in combining such image processing methods with edges/corners detectors in order to benefit from the advantages in both algorithms, similarly to what has been proposed in Rondao and Aouf (2018).

### 2.3.2. POSE ESTIMATION SOLVERS

Several methods exist in the literature to solve for the initial pose of an uncooperative target. Based on two different surveys by Opromolla et al. (2017b) and Sharma and D'Amico (2015), the most commonly used solvers can be identified as the PosIt (Dementhon and Davis, 1995) and Coplanar PosIt (Oberkampf et al., 1996), the SoftPOSIT (David et al., 2004), the EP$n$P (Lepetit et al., 2009) and the Newton-Raphson Method (NRM) (Ostrowsky, 1966). A comparative assessment of the different P$n$P solvers is reported in Table 2.4. In Rondao and Aouf (2018), the EP$n$P solver was used to initialize the pose, which was further refined by means of an M-Estimator minimization to increase the robustness with respect to erroneous correspondences between features. In their method, the Rodrigues parameters were used to represent the relative attitude in order to handle a $6\times1$ pose vector. Sharma, Ventura, et al. (2018) further proved that the EP$n$P method has the highest success rate and offers a superior performance in terms of both pose accuracy and runtime when compared with other state-of-the art P$n$P solvers. In their estimation scheme, the NRM was also used after the EP$n$P to refine the final pose estimation. The idea behind such P$n$P solver switch is that, since EP$n$P has the lowest runtime, it can be used when large number of correspondence hypotheses need to be validated within

the first iterations. Once the search space for correct feature correspondence has been reduced, NRM can be used due to its improved accuracy in the presence of outliers and noise (Sharma and D'Amico, 2015). The same pose estimation system was adapted by Bechini et al. (2022) in order to work better with VIS/TIR fused images. Specifically, edge detection was made less sensitive to variations in images to better estimate the size of the match matrix needed for the pose estimation. Furthermore, Pesce, Opromolla, et al. (2019) proposed a novel pose estimation scheme in which the RANSAC algorithm is used in combination with the Principal Component Analysis (PCA) to generate subsets of image-model correspondences, so called *consensus sets*. For this purpose, the features extracted with the GFTT algorithm were compared with an off-line feature point classification of a simplified 3D model. Once the correspondences are set, the EP$n$P is used to solve for the pose initialization. The SoftPosIt algorithm was further included to solve for the pose tracking. Due to the capability to detect particular spacecraft structures, their estimation scheme proved to be robust with respect to spacecraft symmetry.

Aside from the listed solvers adopted to solve the pose initialization problem, other authors (Galante et al., 2016; Naasz et al., 2009) implemented the technique proposed in Drummond and Cipolla (2002) and the ULTOR engine (Hannah, 2008) in their Goddard Natural Feature Image Recognition (GNFIR) and ULTOR algorithms, respectively, for the pose tracking. As opposed to P$n$P solvers, this technique makes use of the Lie group SO(3) to find and measure the distance between a rendered model of the target and the matching nearby edges in the image. In their works, the GNFIR algorithm was adopted to perform edge tracking once the pose initialization is acquired, whereas ULTOR could be used for both pose initialization and tracking. Additionally, Gansmann et al. (2017) assumed the initialization to be known and implemented a tracking method based on Drummond and Cipolla (2002) which uses an Iteratively Re-Weighted Least Squares (IRLS) to get an a-posteriori pose via the interframe motion. Their algorithm minimized the squared residuals of model template edges, extracted from a 3D rendering of the target, to image query edges, extracted from each TIR image. Their tracking algorithm was tested for the distance from 100 m until 21 m and proved to return centimetric and sub-degree accuracy for the pose. However, convergence to local minima associated to a wrong pose represented an issue with the algorithm. A proposed solution to this problem was to perform a re-initialization of the pose estimation with an acquisition algorithm, as a sudden jump in the estimated pose would be easily detected due to the smoothness of the relative motion. Table 2.5 lists some characteristics of the different pose estimation solvers in relation to feature-based methods that use the image processing algorithms described in Section 2.3.1. From the comparison, the pose estimation scheme proposed by Sharma, Ventura, et al. (2018) stands out as a promising candidate for the pose initialization, given the robustness of its image processing system and the fact that it has been tested for several illumination conditions as well as with the Earth in the background. The proposed system is in fact robust to the background of the images due to the WGE, it requires no a-priori knowledge of the target spacecraft's pose, and it is computationally efficient. In particular, this architecture shows improvements with respect to previous image processing and pose estimation techniques (D'Amico et al., 2014; Opromolla et al., 2017b; Sharma and D'Amico, 2015). Some remarks shall be made about the images used for the validation of the pose estimation methods. As reported in Table 2.4, most of

Table 2.5: Characteristics of state-of-the art model-based pose estimation schemes. Here, NA refers to the fact that no robustness tests could be found in the reference.

| Ref. | Image Processing | Pose Initialization/ Tracking | Tested Range [m] | Robust w.r.t. symmetry | Validation Database |
|---|---|---|---|---|---|
| Naasz et al., 2009 | Digital corr./ Sobel | ULTOR/ GNFIR | 150 - 1 | NA | Flight spare cameras/ Lab pictures |
| Du et al., 2011 | Canny + HT | Analytical | 300 - 1 | NA | Synthetic images Realistic camera model No materials' reflectivity |
| Liu and Hu, 2014 | Ellipses extraction | NRM | 40 - 5 | Yes | Synthetic images Ideal camera model No materials' reflectivity |
| D'Amico et al., 2014 | LPF + Canny + HT | Perceptual Groups + NRM | 13 - 8 | No | Actual space imagery (PRISMA) |
| Shi, Ulrich, Ruel, and Anctil, 2015 | RCM + HCD | SoftPosIt | ~5m | NA | Synthetic images Camera model not given No materials' reflectivity |
| Galante et al., 2016 | Sobel | GNFIR | NA | - | - |
| Shi et al., 2016 | CLAHE + SIFT/BRIEF + RANSAC | EP*n*P/SoftPosit | - | NA | Synthetic and lab TIR images Camera model not given No materials' reflectivity |
| Gansmann et al., 2017 | Canny | IRLS | 100 - 21 | NA | Actual space imagery (ISS) |
| Rondao and Aouf, 2018 | FREAK + EDL | EP*n*P/RANSAC + M-estimator | NA | Yes | Synthetic images Camera model not given Materials' reflectivity included |
| Sharma, Ventura, et al., 2018 | WGE + S/HT | EP*n*P + NRM | 13 - 8 | Yes | Actual space imagery (PRISMA) |
| Pesce, Opromolla, et al., 2019 | GFTT | RANSAC + PCA + EP*n*P/SoftPosIt | < 30 | Yes | - |
| Bechini et al., 2022 | WGE + S/HT | EP*n*P + NRM | < 30 | Yes | Synthetic PRISMA images |

the pose estimation schemes were tested with synthetic images in which the different reflectivities of spacecraft materials were not included. As such, the robustness of the algorithms with respect to realistic illumination conditions and target's textures could not be assessed. Also, the limited amount of realistic space images available in D'Amico et al. (2014), Gansmann et al. (2017) and Sharma, Ventura, et al. (2018) could not represent all the challenging orbital scenarios for which a specific camera-target-Sun-Earth geometry would affect the pose estimation accuracy.

### 2.3.3. POSE ESTIMATION ARCHITECTURES

When the image processing algorithms and pose estimation solvers described in Sec. 2.3.1-2.3.2 are used for the estimation of the pose (feature-based methods), the adopted interface between image processing and pose estimation can result in different system architectures. Figure 2.4 illustrates the three standard options adopted in the literature using PRISMA's Tango spacecraft as a reference target. In a type-A architecture, features detected and extracted from the input image are matched with the features of an existing wireframe model of the target and fed to a pose estimation solver. Since edges and corners

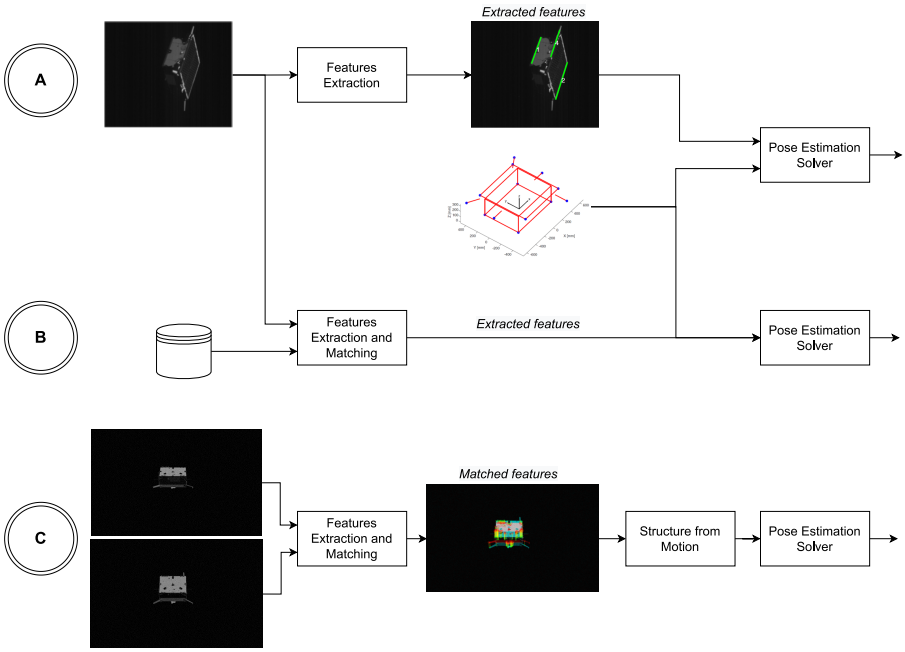



Figure 2.4: Standard pose estimation architectures using PRISMA's Tango images as a reference target.

are more easily matched with features from a simplified model, the Harris corner detector or more generic line/corner detectors are typically implemented at an image processing level. Note that since the type and number of extracted features varies based on the illumination conditions and target orientation, the matching with the wireframe features could result in a lengthy process that could jeopardize the pose estimation. For this reason, in type-B architectures the features extracted in the input image are first matched with an image database. This is done in order to select pre-defined features from the database that can be easily matched with the wireframe model without the need of a large search-space for the feature correspondences. Generally, Keypoint features detectors (i.e. SURF, ORB, SIFT, FREAK) are preferred over edge/corner detectors as they extract texture-rich points on the target that can be matched more easily with the offline images. However, inconsistent matches can be obtained if the input image and the offline images differ considerably in illumination conditions and target textures. Alternatively, type-C architectures avoid the use of an offline database and perform feature extraction and matching between two (or more) subsequent images. The matched features are then used to reconstruct an online model of the target via model reconstruction methods such as Structure from Motion (SfM), before feeding the 2D/3D correspondences to the pose estimation solver. Although some of the challenges involved in the matching process with an offline database are avoided, large inter-frame rotations of the target as well as large variations in the illumination conditions can still result in unfeasible matches between subsequent frames, especially when dealing with actual space imagery. Moreover, the

additional uncertainties involved in the model reconstruction can impact the overall pose estimation accuracy.

### 2.3.4. APPEARANCE-BASED POSE ESTIMATION



Compared to the feature-based methods shown in Figure 2.4, in which the image processing is used to extract features such as keypoints, corners and edges, only the spacecraft *appearance* is used in appearance-based methods. Depending on whether a 3D model of the target spacecraft is used or not, appearance-based methods can be classified as model-based and model-free, respectively. Opromolla et al. (2017a) proposed a model-based pose framework for spacecraft pose estimation. However, the framework was designed to process 3D point clouds and thus its application was constrained to LIDARs or stereovision systems. To the best of the author's knowledge, the only appearance-based method for spacecraft pose estimation based on a monocular camera was proposed by Shi et al. (2017) and is based on PCA. The pose matching algorithm is separated into an off-line training portion and a testing portion that computes the pose of the spacecraft in-flight. The PCA algorithm matches the object from the camera image (test image) to a stored matrix of images that has been transformed to its eigenspaces during the training phase. The advantage of PCA stands in the fact that the dimension of the training dataset can be drastically reduced by considering only the principal eigenvectors of the training data matrix. However, the test image needs to be compared to each image of the training dataset at each pose solution, which still requires a considerable computational effort if the number of training frames is large. In Shi et al. (2017), the validation of the algorithm was performed with $M = 12,660$ frames as a result of a trade-off between the computational time and the estimation accuracy. The resulting mean search time was found to be approximately 62.8 ms, which is relatively low for uncooperative pose estimation. However, the PCA algorithm performance was proven to degrade in noisy images, which is unwanted when actual space imagery is used as input. Furthermore, one of the assumptions for the PCA is that the object must be completely visible, which might not be the case if part of the spacecraft is outside the camera FoV. Finally, as the validation was not performed with the Earth in the background, it is unclear whether the pose estimation is robust against other objects present in the camera image, as one of the main requirements of PCA is that each image shall contain a single, non-occulted object.

### 2.3.5. CNN-BASED POSE ESTIMATION

Recent advances in Computer Vision (CV) for pose estimation in terrestrial applications have relied on the quickly evolving domain of machine learning to enable unprecedented efficacy across several applications. Among deep learning techniques, CNNs stand out as a promising and viable solution for visual-based systems, mostly due to their capability to process visual data in form of images. CNNs are named so because they use convolution operation in at least one layer of the network. The implementation of CNNs for monocular pose estimation in space has already become an attractive solution in recent years, also thanks to the creation of SPEED and SPEED+ datasets (Kisantal et al., 2020; Park, Martens, et al., 2021), which include highly representative synthetic and lab-generated images of PRISMA's Tango spacecraft made publicly available by Stanford's Space Rendezvous Laboratory and applicable to train and test different network architectures.

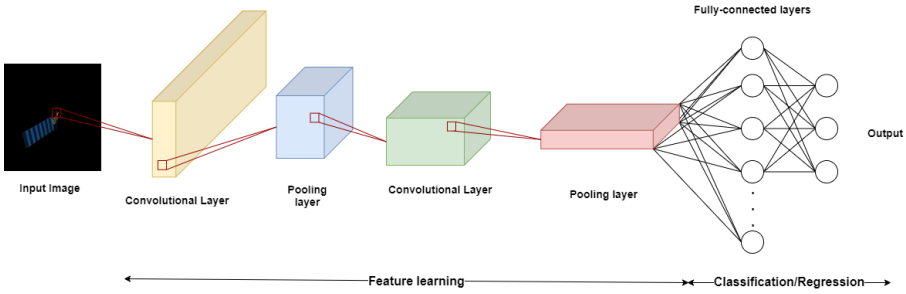



Figure 2.5: High-level schematic of a generic CNN architecture.

A high-level schematic of a typical CNN architecture is shown in Figure 2.5. The first and core building block is represented by the Convolutional layer, formed by a fixed number of weight filters which convolve the input image to return feature maps. Generally, a convolution layer uses multiple weight filters on the image. As such, the output is a 3D volume of 2D feature maps. These feature maps are then down-sampled by the Pooling layer in order to progressively reduce the spatial size of the feature maps, thus reducing the number of parameters and improving the network robustness against overfitting. By cascading Convolutional layers with pooling layers in series, the output of the so-called feature learning step is a list of feature maps of lower size than the input image. Fully-connected layers are then generally used to classify the image given the input feature maps.

The interesting aspect of CNNs is that the internal representations that express the relationship between two feature maps are automatically generated. As already mentioned, CNNs are capable of extracting abstract representations at unintuitive levels of information in the images, to compose high level information. Therefore, CNNs have been successful in tackling some of the challenges faced by image processing systems, where a higher interpretability of the information is required. These include, among others, robustness against:

- viewpoint variation;
- scale variation, deformation, occlusion;
- illumination conditions;
- background clutter.

Altogether, these represent the main advantages of CNNs over standard image processing algorithms for pose estimation (Capuano et al., 2019; Rondao et al., 2018; Sharma, Ventura, et al., 2018). Notably, CNNs for vision tasks are most often trained offline in a supervised manner, wherein training images are fed to the network and the appropriate convolutional and activation weights are estimated to reduce the loss. This means that there is no need to correlate the offline images with the images taken in orbit, as it usually occurs in the feature matching step of standard keypoint-based methods (Section 2.3).

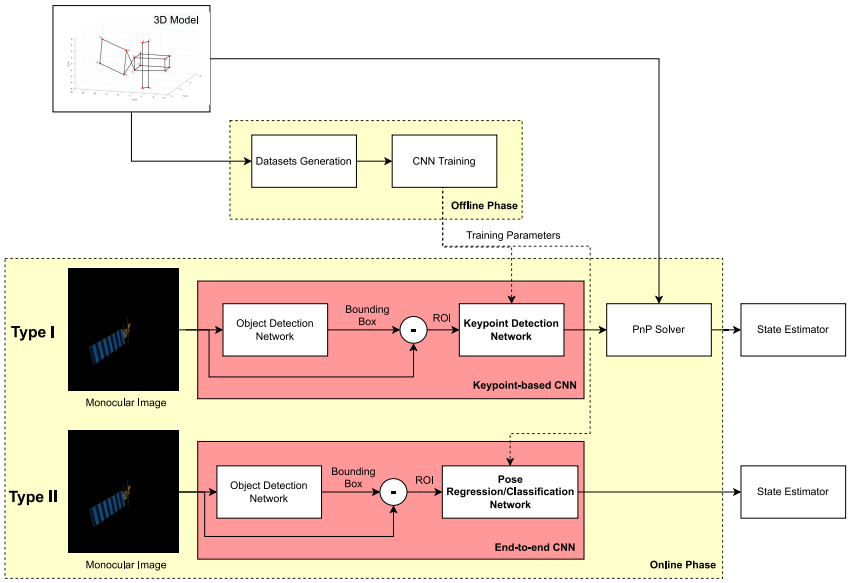



Figure 2.6: Implementation of CNNs for pose estimation. Two system architectures are typically used in which the CNN is either used to detect keypoints (Type I) or to directly infer the pose (Type II).

**Estimation Architectures** From a high-level perspective, there are two possible pose estimation systems based on CNNs. These are compared in Figure 2.6. In each of them, the first essential step is represented by an Object Detection Network (e.g. Faster R-CNN (Ren et al., 2017), R-FCN (Dai et al., 2016) or MobileNet-based (Howard et al., 2017) placed before the main CNN. The CNNs employed in these detection architectures solve two problems:

- regression of bounding box coordinates;
- classification of the object class (when more targets could be in the camera FoV).

This is done in order to detect the target object in the image and crop a Region of Interest (ROI) around it. For monocular pose estimation in space, this step is crucial as ROI cropping allows robustness to scale and background textures. In Type I systems, the cropped ROI is fed into a Keypoint Detection Network. These CNNs keep only the feature learning step of Figure 2.5, and directly output a set of feature maps without classifying the object. These so-called heatmaps are detected around pre-selected features on the target object, such as corners or interest points. The 2D pixel coordinates of the heatmap's peak intensity characterize the predicted feature location, with the intensity and the shape indicating the confidence of locating the corresponding keypoint at this position, i.e. the prediction accuracy (Pavlakos et al., 2017). Notably, the selection of the specific CNN architecture will drive the achievable keypoints detection accuracy and robustness. Some architectures, such as the Hourglass (Newell et al., 2016) and the U-Net (Ronneberger

et al., 2015), perform a down-sampling of the input followed in series by an up-sampling, in order to detect features at different scales. However, recent advancements in the field demonstrated that by using parallel sub-networks across multiple resolutions, rather than multi-resolution serial stages, the CNN can manage to maintain a richer feature representation, facilitating more accurate and precise heatmaps (Chen et al., 2019). In this context, the High-Resolution Net (HRNet) (Sun et al., 2019) currently stands out as a promising architecture for keypoint detection. Once the pre-selected keypoints are detected by the CNN, the extracted keypoints are ultimately fed to a standard P$n$P solver (Section 2.3.2) together with their body coordinates, which are made available through the wireframe 3D model of the target body.

Conversely, Type II systems (also called end-to-end systems) predict the full pose vector directly from the input image, based on either classification or regression following the feature learning step (see Figure 2.6). Classification formulations discretize the pose space and use the fully-connected layers to convert the feature maps into a confidence score for each possible pose vector. On the other hand, the regression formulation aims to learn a non-linear mapping that directly regresses the six pose variables. Pose classification formulation has been utilized in terrestrial applications (Su et al., 2015) as well as for uncooperative spacecraft (Sharma, Beierle, et al., 2018), generally using the AlexNet (Krizhevsky et al., 2012) or VGG16 (Simonyan and Zisserman, 2014) architectures. On the other hand, the regression formulation has been used most often in terrestrial applications (Mahendran et al., 2017; Shi, Ulrich, and Ruel, 2015).

Initially, end-to-end CNNs were more commonly exploited for the pose estimation of uncooperative spacecraft (Sharma, Beierle, et al., 2018; Sharma and D'Amico, 2019; Shi et al., 2018; Sonawani et al., 2020). However, since the pose accuracies of these systems proved to be lower than the accuracies returned by standard P$n$P solvers, especially in the estimation of the relative attitude, systems based on keypoints-detection recently emerged as the preferred option. Specifically, average orientation errors of $1.31° \pm 2.24°$ were achieved by keypoint-based methods as opposed to the average orientation errors of $9.76° \pm 18.51°$ achieved by end-to-end methods. These averages were computed across test images of the Tango spacecraft as part of the SPEED challenge (Huo et al., 2020; Kisantal et al., 2020). Notably, another advantage of keypoint-based methods resides in their capability to return both the 2D location of the detected features (CNN output) and the full pose (P$n$P output). Consequently, they can provide a much more flexible interface with the navigation filter, as will be discussed in Section 2.4.3.

**The domain shift problem** Although CNN-based systems already proved to perform better than standard image processing algorithms under several scenarios with adverse orbital conditions (e.g. image noise, near-eclipse conditions and adverse target reflections), their performance on actual space imagery is still challenged by the fact that their layers are trained on purely synthetic images, which typically fail at representing realistic space conditions and target textures. Referred to as the domain shift problem in Chapter 1, this aspect can undermine the CNN performance when the target domain (i.e. the real space environment) differs from the training domain (i.e. the simulated synthetic

environment), and is a direct result of a lack of large training datasets of representative images of space objects. In this context, it becomes paramount to improve the performance of synthetically-trained, CNN-based pose estimation systems on realistic imagery that mimic actual space imagery. Various works have been carried out in recent years to leverage the domain shift from synthetic training to real test imagery, either via *data augmentation* (Geirhos et al., 2019; Jackson et al., 2018; Tobin et al., 2017) or via *domain adaptation* (Donahue et al., 2017; Ghifary et al., 2016). Although domain adaptation techniques are often effective and can produce impressive results by adapting the CNN on a specific target domain post training, they require the target domain images and synthetic training images simultaneously to perform adaptation, and hence they are not domain-agnostic. On the other hand, data augmentation techniques consist of introducing variations in the synthetic training domain without any a-priori knowledge of the target domain. In essence, the idea is to extend the standard data augmentation effects, such as random cropping, zooming, rotation, flipping etc. with texture and complex illumination variations. By doing that, Tobin et al. (2017) already showed that a CNN can generalize from synthetic environments to new domains by using an unrealistic but diverse set of random textures. Following this line of reasoning, other authors further discovered that by randomizing textures during training, CNNs can learn the shape of objects rather than textures, improving their robustness to domain shift (Geirhos et al., 2019; Jackson et al., 2018). However, these methods were tested on terrestrial applications, and their applicability to pose estimation in space has not been fully proven yet. As such, further research is needed in order to bridge this performance gap. In essence, the domain shift problem and the validation on actual space imagery are fundamental aspects which shall be accounted for in any CNN-based pose estimation systems.

## 2.4. VISUAL-BASED NAVIGATION FILTERS

In the framework of spacecraft relative motion, several representations of a linearized relative state (position, attitude and translational/rotational velocities) exist based on the intersatellite range, orbital eccentricity and perturbation forces involved. These representations are used to describe the relative motion of a target object with respect to a servicer spacecraft in a co-moving Local Vertical Local Horizontal (LVLH) reference frame (Figure 2.7). Linearized models are required when the filter internal dynamics needs to be linearized, as it is the case for linear Kalman Filter (KF) and Extended Kalman Filter (EKF). Sullivan et al. (2017) provide a detailed overview on closed-form dynamics model suited for onboard relative navigation. Notice that, for ADR and OOS, the target orbit can usually be assumed to be circular, thus simplifying the computational burden that results from not neglecting the orbital eccentricity of satellite orbits. Generally, a distinction is made between models which make use of a Cartesian representation of the relative state (position and velocity) and models which consider a set of the Relative Orbital Elements (ROE). Notably, perturbation models can be easily accommodated in the filter dynamics in the latter case (Guffanti et al., 2017; Hamel and de Lafontaine, 2007; Koenig et al., 2017). Clearly, a linearized model is not required if nonlinear filters are adopted. On the other hand, in the context of spacecraft relative attitude, several linear and nonlinear models exist based on either Euler angles, quaternions and Modified Rodrigues Parameters (MRP)

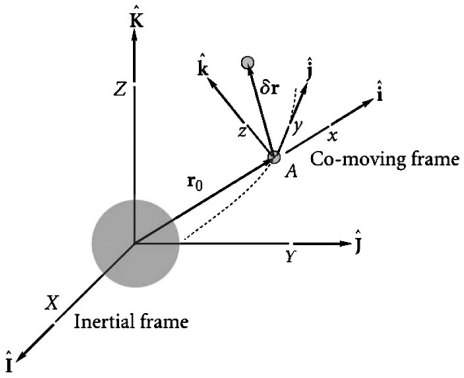



Figure 2.7: Co-moving LVLH frame (Curtis, 2005).

(Kim et al., 2007; Lefferts et al., 1982; Markley, 2003).

Navigation systems for close-proximity operations have been extensively validated in the context of RF and monocular vision-based navigation for OOS as well as other rendezvous missions, when the target is cooperative (Allende-Alba et al., 2009; Branco et al., 2015; D'Amico et al., 2013; Kim et al., 2007; Zhang et al., 2015). However, there is still a lack of a comprehensive validation of navigation systems for the pose estimation of an uncooperative target. As an example, the EKF and the Unscented Kalman Filter (UKF) presented in Kim et al. (2007) and Zhang et al. (2015), respectively, rely on the availability of gyro measurements from each spacecraft, which is usually not the case for uncooperative spacecraft in ADR scenarios. When the uncooperative target is known, it is assumed that a simplified geometrical model of the target is available and representative of the target state in orbit. As such, when a model-based pose estimation method is adopted prior to the navigation filter, the 3D model of the target can be assumed to be reliable, and the navigation system can estimate the pose based on the pseudomeasurements derived from the extracted features of the target without including uncertainty in the geometrical model. However, if the shape of the target has changed due to orbit degradation or due to unforeseen events, the assumptions on its state made in the simplified geometrical model might differ from its real conditions in orbit. Furthermore, the target's mass and moment of inertia, together with other relevant parameters, might differ from the assumed values. As such, the navigation filter might have to estimate additional parameters aside from the pose.

### 2.4.1. DESIGN AND VALIDATION OF MONOCULAR NAVIGATION SYSTEMS: KNOWN TARGETS

When dealing with uncooperative known targets, the state vector to be estimated in the navigation filter consists in the relative position, velocity, attitude and angular velocity between the servicer and the target. Additionally, if the relative dynamics modeled in the relative navigation system account for perturbation models which might be inaccurate, key perturbation parameters should be included given the uncertainty of the dynamics

Table 2.6: Comparison of navigation filters for pose estimation, together with the adopted performance validation method. Here, NS refers to papers in which the adopted filters were not specified

| Ref. | Translational filter | Rotational filter | Performance Validation Method |
|---|---|---|---|
| Naasz et al., 2009 | Linear KF | MEKF | Ground-based test on HST mockup |
| Sharma and D'Amico, 2017 | MEKF | MEKF | Numerical simulations |
| Gasbarri et al., 2014 | Linear KF | Linear KF | HIL in closed GNC loop |
| Galante et al., 2016 | MEKF/ Schmidt KF | MEKF/ Schmidt KF | Numerical simulations |
| Filipe et al., 2015 | D-Q MEKF | D-Q MEKF | Ground-based experimental test |
| Colmenarejo et al., 2019 | NS | NS | SIL/HIL in closed GNC loop |
| Cavenago et al., 2018 | DA filters | DA filters | Numerical simulations |
| Pesce, Haydar, et al., 2019 | - | Minimum Energy Filter<br>Attitude Observer<br>2nd Order Minimum Energy Filter<br>MEKF | Numerical simulations |
| Pesce, Opromolla, et al., 2019 | $H_\infty$ filter | 2nd Order Minimum Energy Filter | Numerical simulations |
| Park and D'Amico, 2022a | Adaptive UKF | Adaptive UKF | Numerical simulations<br>HIL images |

models. As already mentioned, loosely-coupled navigation architectures are usually preferred when the target is known.

Table 2.6 lists the state-of-the-art for the navigation filters adopted in the framework of pose estimation of uncooperative known targets. Naasz et al. (2009) implemented a Multiplicative Extended Kalman Filter (MEKF) (Markley, 2003) for attitude estimation and a linear KF for translation to estimate the pose of the HST, assumed to be uncooperative. Furthermore, Sharma and D'Amico (2017) proposed a reduced-dynamics pose estimation in which a MEKF is formulated, validated and stress-tested with the PRISMA dataset. The measurement model was computed from pseudomeasurements, derived from the line segments detected from the image by the image processing, by expressing each line segment as a function of the ROE and of the relative attitude quaternion. However, in both implementations the filter dynamics were highly simplified and no perturbation models were included. Moreover, the initial conditions for the relative state in Sharma and D'Amico (2017) were assumed from the separate results of the pose initialization subsystem, without modeling the interface between the initial pose estimation and the filter itself, and no Software-In-the-Loop (SIL)/Hardware-In-the-Loop (HIL) tests were conducted. Gasbarri et al. (2014) performed a HIL experiment in a closed GNC loop using the camera as a standalone sensor. However, no perturbation models were included in the filter dynamics and only a simplified linear KF was implemented. Galante et al. (2016) proposed the fusion of several measurements from different types of monocular sensors and a LIDAR in a MEKF. Their navigation filter was designed assuming that no information about the servicer absolute position and velocity is available. As such, they

neglected orbital dynamics in the filter propagation step, and considered a Schmidt KF (Schmidt, 1966) to counteract the limited system observability, which results from the lack of sufficient richness in the relative motion dynamics. Furthermore, the filter state was augmented with sensor biases to account for the different optical spectra of the pose measurement sensors. Filipe et al. (2015) validated experimentally a Dual Quaternion MEKF (DQ-MEKF) (Markley, 2003) suitable for uncooperative satellite proximity operation scenarios, in which the pose measurements are rearranged in a dual quaternion form and fed into the navigation filter. Their filter proved to be fast enough for operational use and insensitive to singularity problems, due to its error formulation. However, only limited scenarios were simulated in the tests. Colmenarejo et al. (2019) performed a comprehensive ground testing to investigate system, as well as subsystems, level considerations related to several ADR scenarios. A complete GNC model designed in a FES was SIL/HIL-tested, thus accounting for the interfaces between the navigation filter, the image processing and the initial pose estimator. Results validated several aspects of the filter robustness, such as information about the illumination quality and sensitivity to blackouts. However, several challenges behind fusing different absolute and relative sensors in the navigation filter were not solved, and the robustness of the navigation filter was not fully investigated. Furthermore, the testing did not account for recent image processing methods, and the robustness of the filter with respect to a tumbling scenario was not assessed. Cavenago et al. (2018) proposed two innovative nonlinear filters based on Differential Algebra (DA) to limit the computational time while preserving the filter performance. Their design included relative rotational dynamics which account for the apparent torques, the servicer-inertial torques and the target inertia matrix, thus improving other models which assumed simplified, unperturbed relative rotational motion. However, only a simplified software was used for the validation of the navigation system. Pesce, Haydar, et al. (2019) decoupled the translational and rotational motion, and compared nonlinear filtering techniques to a MEKF for the relative attitude estimation of an uncooperative target. Nonlinear filtering algorithms such as the Minimum Energy Filter, the Attitude Observer (Mortensen, 1968; Zamani et al., 2013), and the 2nd Order Minimum Energy Filter (Zamani et al., 2014) were adapted for the specific application. Compared to the analysis conducted by Cavenago et al. (2018), the filters performance was assessed by considering limited knowledge on the target inertia matrix by neglecting the relative dynamics in their formulation. Their results showed that, despite a quicker convergence in transient, the MEKF has a lower performance at steady-state when compared to the nonlinear filters. Furthermore, the second-order minimum energy filter without dynamics was proposed as the best option in scenarios where neither the angular velocity nor the inertia matrix of the target are fully known. Furthermore, Pesce, Opromolla, et al. (2019) proposed a novel navigation system in which a $H_\infty$ Filter (Simon, 2006) was selected for the translational motion estimation and the 2nd Order Minimum Energy Filter for the rotation motion estimation, respectively. The translational filter implemented the Yamanaka-Ankersen (Yamanaka and Ankersen, 2002) formulation of satellite relative motion, and it was chosen based on the claim that assumptions of KF are usually not satisfied when dealing with optical systems, and on the fact that the absolute position of the servicer, together with the illumination conditions, can strongly affect the process and measurement noise if a KF is selected. Their design returned

a navigation system for which filter robustness is preferred rather than filter accuracy. On the other hand, the selected rotation filter was characterized by a null derivative of the angular acceleration, in order to avoid the dependence of the filter accuracy on the knowledge of the inertia matrix of the target spacecraft. Despite the worse performance compared to filters that include the relative dynamics, and thus the inertia matrix of the target, the proposed formulation could be extended for the pose estimation of partially known targets. Results obtained by considering LEO, Highly Elliptical Orbit (HEO) and Geostationary Earth Orbit (GEO) scenarios showed a steady state relative position and attitude Root-Mean-Square Error (RMSE) lower than 3 cm (except for HEO) and 1 degree, respectively. Notice also that no perturbation models were included in both filters. In a recent effort to extend the validation of relative navigation filters, Park and D'Amico (2022a) proposed an adaptive UKF and performed a validation of the navigation filter with both synthetic images and HIL images of PRISMA's Tango spacecraft. By including an adaptive scheme to compensate for the uncertainty in the process noise covariance, the filter was proven robust towards noisy measurements and large uncertainties in the filter dynamics.

An important aspect of the relative navigation filter reviewed so far relates to whether the absolute state of the servicer spacecraft is required to estimate the relative state between the servicer and the target. Except for the design in Galante et al. (2016), the reviewed filter designs assumed that the absolute state of the servicer is known, which implies that absolute sensors such as GPS and Inertia Measurement Units (IMU) shall be included in the absolute filter. However, GPS can increase complexity to the system and it is not being considered in some of the current designs for close-proximity rendezvous missions. On the other hand, the limited accuracy of the absolute pose and velocity information from an Inertial Measurement Unit (IMU) onboard the servicer would probably result in a decreased accuracy in the estimated relative state, when compared to the estimation accuracy results obtained by assuming no noise in the absolute position and velocity. It can be stated that the interface between the relative and absolute navigation filters onboard the servicer spacecraft still presents open issues.

### 2.4.2. DESIGN AND VALIDATION OF MONOCULAR NAVIGATION SYSTEMS: PARTIALLY KNOWN TARGETS

During close-range rendezvous, the relative attitude dynamics strongly depends on the target's moment of inertia, which might be partially unknown for inactive satellites. At the same time, the knowledge of the location of the center of mass is critical for a safe approach to the target. As such, it is important to include the estimation of these parameters in the navigation filter, in order to improve the knowledge of the target state as well as of its orbit relative to the servicer. The position and velocity of the center of mass can be estimated by solving a least squares problem in which the position and velocity of the geometrical center, or of a feature point, on the target body are measured by a monocular camera (Benninghoff and Boge, 2015; Sheinfeld and Rock, 2009). Alternatively, Al-Isawi and Sasiadek (2018) calculated the location of the center of mass using kinematic equations and an Iterative Closest Point (ICP) algorithm, and Meng et al. (2018) implemented an EKF and additionally estimated the target body mass by applying an impulse to the

target.

Several approaches exist in literature to estimate the target moment of inertia with stereo cameras or other active sensors such as LIDARs. The interested readers are referred to the survey conducted by Opromolla et al. (2017b) for a comprehensive overview. However, there are more restrictions on system observability when monocular cameras are adopted. Sheinfeld and Rock (2009) presented a framework for rigid body inertia estimation for torque-free and non torque-free motion applicable to monocular vision. Following these findings, Benninghoff and Boge (2015) and Qiu et al. (2017) proposed two methods based on kinematic equations and the conservation of angular momentum, in combination with a constrained least squares method, to ensure positive diagonal values of the inertia matrix. Additionally, Hou et al. (2017) proposed a dual vector quaternions-based EKF and a dual vector quaternions-based adaptive fading factors EKF to estimate the ratios of the inertia parameters of a free-floating tumbling space target. In all these methods, only normalized moments of inertia were estimated, since no external torques were applied on the target spacecraft. Setterfield et al. (2018) proposed a method to additionally estimate the three principal axes together with the inertia ratios through the analysis of the target object's polhode in an arbitrary target-fixed geometric frame. Felicetti et al. (2014) analyzed the estimation of the full inertia matrix by exerting a control torque on the object and by adopting an EKF. However, their method is applicable only to estimate the moment inertia of the multibody system once the chasing and the grasping phases have occurred. Xu and Wang (2017) investigated the possibility to estimate the target inertia by using the information of the mass and velocity of a bullet shot to the target to change its angular momentum. Recently, Meng et al. (2018) proposed a different method based on the application of a number of impulses to the target in order to observe the resulting motion changes and solve for all the inertia parameters. An EKF was used to estimate the normalized inertia matrix together with the target mass, and a least squares method was added to estimate the full set of inertial parameters.

### 2.4.3. CNN-BASED NAVIGATION FILTERS

Beside the specific aspects within a navigation filter, additional challenges arise in relation to the applicability of CNNs for relative navigation in space. In fact, limited focus has been given to the interface between a CNN-based system and a navigation filter, despite the recent advances in navigation techniques.

As already discussed in Section 2.3.5, one of the major differences between end-to-end and keypoint-based pose estimation systems resides in the intermediate estimates that can be output at each estimation step. In a keypoint-based system (Type I in Figure 2.6), both keypoints location and pose can be estimated by cascading a CNN-based keypoints detector with a standard P$n$P solver. Conversely, only an estimate of the pose can be output by an end-to-end system (Type II). As a result, the selection of the pose estimation system will drive the selection of the type of navigation filter. This is shown in Figure 2.8.

In a tightly-coupled filter, the measurement error covariance matrix is constructed with the covariances of each detected feature. If a keypoints regression architecture

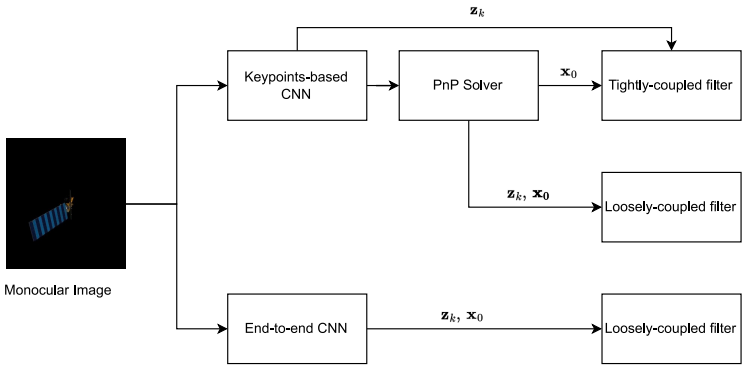



Figure 2.8: Overview of the three different interfaces between the CNN, the P$n$P solver and the navigation filter. Here, $\mathbf{x}_0$ and $\mathbf{z}_k$ represent the initial state vector and the measurements vector, respectively.

(Type I) is chosen at a pose estimation level, a covariance matrix will be associated to each feature detected by the CNN and could, theoretically, be derived from the CNN heatmaps. Conversely, in the loosely-coupled filter the measurement error covariance represents the uncertainty in the pose estimation step, and hence it cannot be directly related to the CNN output. In case a keypoint-based CNN is exploited together with a standard P$n$P solver (Type I), no statistical information can be easily retrieved from the pose estimation, and hence a constant covariance is usually chosen based on the pose estimation accuracy observed for the validation or test datasets. If on the other hand an end-to-end CNN is used (Type II), the single CNN confidence returned for each estimated pose cannot be easily associated to each element of the relative position and attitude. To compensate for this lack of statistical information, recent implementations (Black et al., 2021b) accommodated an error prediction network after the main end-to-end CNN which rejects wrong pose estimates. As such, inaccurate measurements can be rejected prior to the navigation step and a constant measurements matrix can be used without undermining filter robustness.

## 2.5. CHAPTER CONCLUSIONS

This Chapter presented a detailed review of the robustness and applicability of state-of-the-art monocular pose estimation systems for the relative navigation with an uncooperative spacecraft. This review is motivated by the applicability of pose estimation in future space missions, i.e. ADR and OOS, which involve close-proximity operations of a servicer spacecraft around a target object. Monocular systems are analyzed due to the strict power, mass, and operational range requirements driving the current design of these missions, which are usually key requirements for active, as well as stereo, systems.

The challenges involved in monocular-based systems are traced back to each individual component of the pose estimation process, from the image processing algorithms to the navigation filter architectures. The key findings are as follows:

1. Due to the limited robustness of VIS/NIR cameras against adverse orbital conditions

and the limited image quality which characterizes TIR cameras, multispectral systems represent a potential solution capable of increasing the overall system robustness, while at the same time preserving system accuracy.

2. Although feature synthesis schemes are able to combine the advantages of standard feature detectors into a more robust image processing system, it is unclear whether this system would perform with high accuracy and robustness against adverse scenarios, such as near-eclipse conditions, unfavourable target reflections and Earth in the background.

3. Standard pose estimation architectures present several challenges related to their associated image processing tasks:

   (a) The detection of variable edges and corners on the target object due to changing view conditions can result in a lengthy 2D/3D correspondence and affect the pose estimation accuracy and robustness.

   (b) Pose estimation systems based on offline databases are likely to decrease their performance when features detected in actual space imagery have to be matched with the synthetic offline features.

   (c) The extraction and matching of features across subsequent images can be jeopardized by large inter-frame rotations of the target as well as large variations in illumination conditions.

4. CNN-based pose estimation systems are emerging as a viable alternative to more standard methods, due to their proven robustness towards several adverse orbital conditions. In this context, keypoint-based CNNs are currently being preferred over end-to-end CNNs due to a better pose estimation performance in representative scenarios.

5. A comparison between different existing navigation filters shows that filter selection for the pose estimation of an uncooperative target is, from a high-level perspective, driven by a trade-off between filter robustness and filter optimality. In particular, filter robustness is highly affected by the challenges in representing the measurements uncertainty of monocular sensors, due to highly varying orbital conditions.

Overall, the most important aspect that stands out for monocular-based pose estimation systems is represented by a dire need to establish a validation framework on-ground with representative images as a surrogate of actual space imagery. In CNN-based systems, this need is related to the domain shift problem, for which synthetically-trained CNNs tend to decrease their performance on realistic imagery. To cope with these limitations, data augmentation or domain adaptation methods shall be assessed and extensively tested on large datasets. In this context, the challenges involved in recreating representative TIR images on-ground ad the fact that real TIR images clearly differ from synthetic images suggest that a VIS-only pose estimation system should be evaluated and validated first, despite the promising applicability of a multispectral system.

# 3

# Evaluation of Tightly- and Loosely-coupled Approaches in CNN-based Pose Estimation Systems

## 3.1. Introduction

Building on the findings reported in Chapter 2, this Chapter introduces a CNN-based pose estimation system and investigates the challenges of interfacing a keypoint-based CNN with both a pose estimation solver and a navigation filter. This work stems from the performance challenges of standard image processing algorithms when tested on actual space imagery or lab-generated imagery, with adverse illumination conditions and large inter-frame rotations of the target. The selection of a keypoint-based CNN at image processing level follows the recently identified benefits of keypoint-based methods over end-to-end methods (Kisantal et al., 2020). The proposed CNN replaces image processing by identifying a set of pre-selected features on the target spacecraft, which are fed to a P$n$P solver prior to the navigation filter (loosely-coupled) or directly to the navigation filter as measurements (tightly-coupled).

Standard pose estimation solvers such as the Efficient Perspective-n-Point (EP$n$P) (Lepetit et al., 2009), the Efficient Procrustes Perspective-n-Point (EPP$n$P) (Ferraz et al., 2014b), or the multi-dimensional Newton Raphson Method (NRM) (Ostrowsky, 1966) do not have the capability to include the uncertainties of the detected features. Only recently, the Maximum-Likelihood P$n$P (MLP$n$P) (Urban et al., 2016) and the Covariant EPP$n$P (CEPP$n$P) (Ferraz et al., 2014a) solvers were introduced to exploit statistical information by including feature covariances in the pose estimation. Ferraz et al. (2014a) proposed a

Parts of this chapter have been published in Pasqualetto Cassinis, Fonod, Gill, Ahrns, and Gil-Fernandez (2021).

method for computing the covariance which takes different camera poses to create a fictitious distribution around each detected keypoint. Other authors proposed an improved pose estimation method based on projection vector, in which the covariance is associated to the image gradient magnitude and direction at each feature location (Cui et al., 2019), or a method in which covariance information is derived for each feature based on feature's visibility and robustness against illumination changes (Harvard et al., 2020). However, in all these methods the derivation of features covariance matrices is a lengthy process which generally cannot be directly related to the actual detection uncertainty. Moreover, this procedure could not be easily applied if CNNs are used in the feature detection step, due to the difficulty to associate statistical meaning to the image processing tasks performed within the network. In this context, another procedure could be followed in which the output of the CNNs is directly exploited to return relevant statistical information about the detection step. This could, in turn, provide a reliable representation of the detection uncertainty.

To the best of the author's knowledge, the reviewed implementations of CNNs feed solely the heatmap's peak location into the pose estimation solver, despite multiple information could be extracted from the detected heatmaps. Only in Pavlakos et al. (2017), the pose estimation is solved by assigning weights to each feature based on their heatmap's peak intensities, in order to penalize inaccurate detections. Yet, there is another aspect related to the heatmaps which has not been considered. It is in fact hardly acknowledged how the overall shape of the detected heatmaps returned by CNN can be translated into a statistical distribution around the peak, allowing reliable feature covariances and, in turn, a robust navigation performance. Deriving an accurate representation of the measurements uncertainty from feature heatmaps can in fact not only improve the pose estimation, but can also benefit the estimation of the full relative state vector, which would include the pose as well as the relative translational and rotational velocities.

Remarkably, the adoption of a CNN in the feature detection step can overcome the challenges in feature tracking by guaranteeing the detection of a constant, pre-defined set of features. At the same time, the CNN heatmaps can be used to derive a measurement error covariance matrix and improve filter robustness at a navigation filter level. Following this line of reasoning, a tightly-coupled filter is expected to interface well with a CNN-based system and to outperform its loosely-coupled counterpart. In this framework, the objective of this Chapter is to introduce a novel heatmaps-based representation of feature uncertainty and to combine a CNN-based feature detector with a covariant-based P$n$P solver while evaluating the performance of a proposed tightly-coupled navigation filter against the performance of a loosely-coupled filter. Specifically, the novelty of this work stands in linking the current research on CNN-based feature detection, covariant-based P$n$P solvers, and navigation filters. The main contributions in this Chapter are:

1. To assess the feasibility of a simplified CNN for feature detection.

2. To improve the pose estimation with heatmaps-derived covariance matrices.

3. To compare the performance of tightly- and loosely-coupled navigation filters.

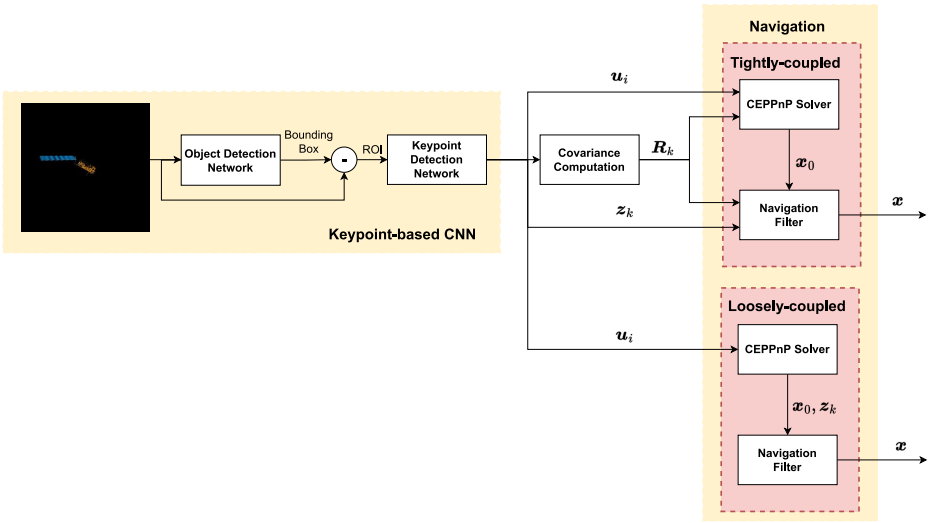



Figure 3.1: Functional flow of the proposed tightly-coupled and loosely-coupled architectures. Here, the detected keypoints $\boldsymbol{u}_i$ are either directly used as measurements $\boldsymbol{z}_k$ in the navigation filter (tightly-coupled) or used by the CEPP$n$P solver to generate pseudomeasurements of the pose (loosely-coupled).

The Chapter is organized as follows. The overall pose estimation framework is illustrated in Section 3.2. Section 3.3 introduces the proposed CNN architecture together with the adopted training, validation, and testing datasets. In Section 3.4, special focus is given to the derivation of covariance matrices from the CNN heatmaps, whereas Section 3.5 describes the CEPP$n$P solver and presents the pose estimation results. Besides, Section 3.6 provides a description of the tightly- and loosely-coupled filters adopted. The simulation environment is presented in Section 3.7 together with the simulation results. Finally, Section 3.8 provides the main conclusions and recommendations.

## 3.2. POSE ESTIMATION FRAMEWORK

Figure 4.11 introduces the pose estimation framework adopted for the evaluation of the two proposed navigation filter architectures. In the keypoint-based CNN block, a CNN is used to extract features $\boldsymbol{u}_i$ from a 2D image of the target spacecraft. Then, statistical information is derived by computing a covariance matrix for each feature, using the information included in the output heatmaps. In the navigation block, a navigation filter estimates the pose as well as the relative translational and rotational velocities. In the tightly-coupled architecture, both the peak locations and the covariances are fed to the navigation filter, which is initialized by a covariant-based P$n$P solver by taking the peak location and covariance matrix of each feature as input and returning the initial pose. In this work, the CEPP$n$P solver is used do to its capability to incorporate feature uncertainty in the estimation. Conversely, in the loosely-coupled architecture the CEPP$n$P directly processes the detected features to output *pseudomeasurements* of the pose, which are fed to the navigation filter at each time step. Notice that thanks to the availability of a

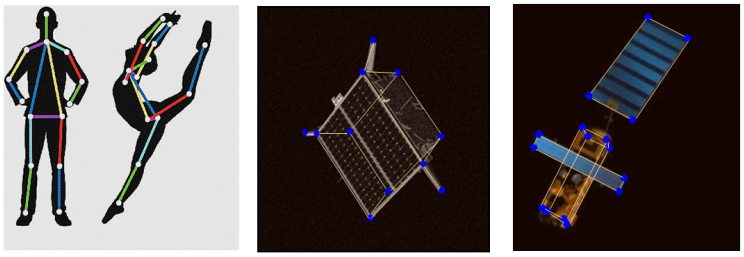



Figure 3.2: Visualization of Keypoint and Skeleton representation for (a) Humans (b) Tango spacecraft (c) Envisat spacecraft (Barad, 2020).

covariance matrix of the detected features, the measurements uncertainty in the tightly-coupled architecture is directly linked to the uncertainty in the CNN detection. This is expected to return a more representative characterization of the measurements, especially in case of inaccurate detection of the CNN due to adverse illumination conditions and/or unfavourable relative geometries between servicer and target. Together with the CEPP$n$P initialization, this aspect can return a robust and accurate estimation of the pose and velocities and assure a safe approach of the target spacecraft.

In this work, a rectilinear vbar approach of the servicer spacecraft towards the target spacecraft is considered, as this typically occurs during the final stages of close-proximity operations in rendezvous and docking missions (Tatsch et al., 2006; Wieser et al., 2015). This assumption is justified by the fact that the proposed method needs to be first validated on simplified relative trajectories before assessing its feasibility under more complex scenarios. Following the same line of reasoning, the relative attitude is also simplified by considering a perturbation-free rotational dynamics between the servicer and the target. These assumptions are described in more detail in Section 3.6.

## 3.3. KEYPOINT-BASED CONVOLUTIONAL NEURAL NETWORKS

Keypoint detection is the computer vision task of localizing and identifying the points of interest, or keypoints, in an image. In this context, keypoint-based CNNs are currently emerging as a promising alternative to more standard feature extraction methods, mostly due to the capability of their convolutional layers to extract high-level features of objects with improved robustness against image noise and illumination conditions. Recently, CNNs have become a popular alternative for challenging detection tasks like Human Pose Estimation (HPE), in which essential points (e.g. knees, elbows and shoulders) on a generic human body need to be detected from an image. For HPE, the human pose is a representation of relative positions of the body keypoints (joints) with a simplistic skeleton (links) that finds its application in activity recognition, robot motion learning, sports injury prediction, and automated sign-language translation among others (Barad, 2020). The success of CNN models stems from their ability to generalize the detection of keypoints on humans of varying sizes and in presence of variations in color and other

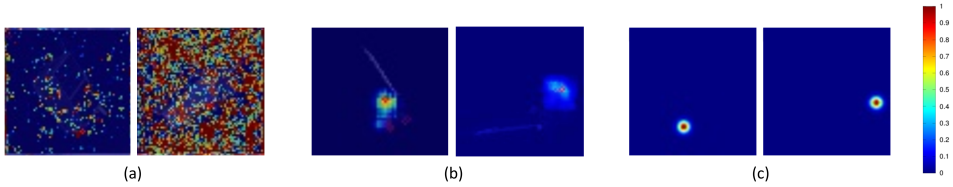



Figure 3.3: Visualization of keypoint prediction heatmaps for (a) bad detections, (b) inaccurate detections, and (c) precise detections (Barad, 2020).



artefacts. As such, they recently stood up as a promising method to tackle the problem of keypoints detection for robust spacecraft pose estimation. Figure 3.2 visualizes detected keypoints in humans and spacecraft use-cases.

In the proposed pose estimation framework, the problem of keypoint detection consists of predicting the coordinates of $k$ predefined keypoints on the spacecraft surface from a 2D image of size $w_k \times h_k$. A quality of modern keypoint-based neural networks is that they also learn to identify keypoints that are not explicitly visible, by learning inherent spatial representation from the provided training examples (Golda et al., 2019). To estimate keypoint coordinates, the state-of-the-art CNNs use a confidence map representation which provides pixel-wise pseudo-likelihood of the true keypoint location being around the detected keypoint. Such a map is also called a *heatmap*. Figure 3.3 shows examples of heatmap outputs from a keypoint-based CNN. A heatmap representation adds keypoint position ambiguity that improves generalization of keypoint predictions, improving generalization and adaptability of the CNN models, in contrast to the direct regression of keypoint coordinates (Szegedy et al., 2016).

### 3.3.1. NETWORK ARCHITECTURE SELECTION

The main challenge associated with keypoint detection using heatmaps lies in the need to recover the spatial resolution required for the output heatmap. Since most CNNs consistently reduce the resolution of the feature maps to infer patterns in the image at the lower levels, the resolution recovery is a crucial part of the detection performance. So far, early state-of-the-art architectures such as the Hourglass (Pavlakos et al., 2017) and the U-net (Ronneberger et al., 2015) decreased the resolution and subsequently recovered resolution through upsampling. Conversely, Sun et al. (2019) recently proposed a radically new architecture that emphasizes on maintaining high resolution throughout the network, called High-Resolution Net (HRNet). By maintaining a rich feature representation which improves the heatmap precision, the HRNet architecture currently represents the state-of-the-art in keypoints detection and had already been successfully tested on Stanford's SPEED dataset (Chen et al., 2019; Kisantal et al., 2020), highlighting its promising application in the pose estimation of a spacecraft.

In this Section, a reduced version of the Hourglass network is selected for the keypoint detection task, and its performance is compared against the state-of-the-art HRNet. This

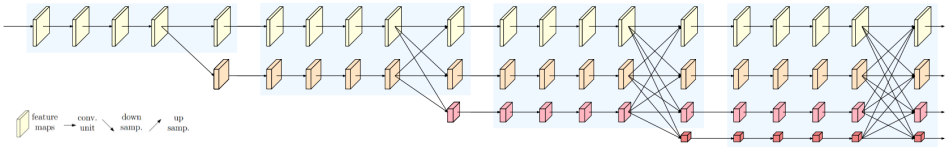
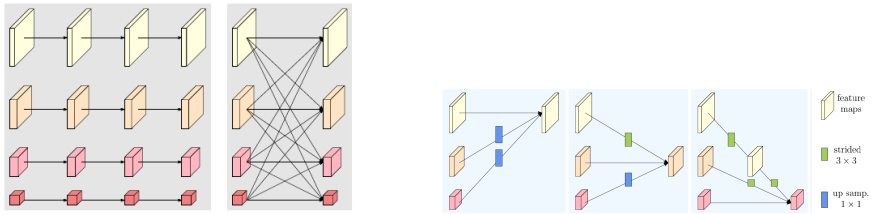



Figure 3.4: HRNet main body architecture with four parallel resolution branches and multi-resolution fusion across branches of feature maps (Sun et al., 2019).






(a) (b)

Figure 3.5: HRNet Visualization: (a) blocks with parallel convolution (residual) units and multi-resolution fusion (b) representation of exchange units used for fusing multiple resolutions.

decision reflects the aim of this research to assess the applicability of less complex and lighter networks (i.e. Hourglass) for keypoint detection in space.

### HRNET

As visualized in Figure 3.4, the HRNet architecture features sub-networks aligned parallel to each other with respective resolution streams. Each parallel network operates on feature maps of the same resolution, and feature maps are repeatedly fused with higher and lower resolution streams between the parallel sub-networks. These elements allow HRNet to maintain high spatial accuracy of the feature maps throughout the network and result in heatmaps that allow a more accurate keypoint detection. The architecture is composed of sub-networks, both parallel and serial. A group of parallel sub-networks across resolution streams is called a *stage*, and the group of serial sub-networks along a resolution stream a *branch*. Then, the network structure of HRNet is constituted by adding one lower resolution sub-network on the subsequent stage.

Each stage consists of blocks (shaded blue in Figure 3.4) containing a group of parallel convolutions called residual units, and a multi-resolution fusion unit as shown in Figure 3.5. Each parallel residual unit is composed of two sets, each comprising a $3 \times 3$ convolution layer, a batch normalization layer and an activation (ReLU) layer. The multi resolution fusion is done by using: (a) a $3 \times 3$ convolution for the feature map at the same resolution, (b) a 2-strided $3 \times 3$ convolution for down-sampling the feature map at the higher resolution, and (c) a $1 \times 1$ convolution followed by a nearest-neighbour up-sampling. Furthermore, in each downward branch the resolution is halved and the channels (convolution kernels) are doubled. Finally, the output heatmaps are obtained from the highest resolution branch of the last stage at the end of the last block, using a

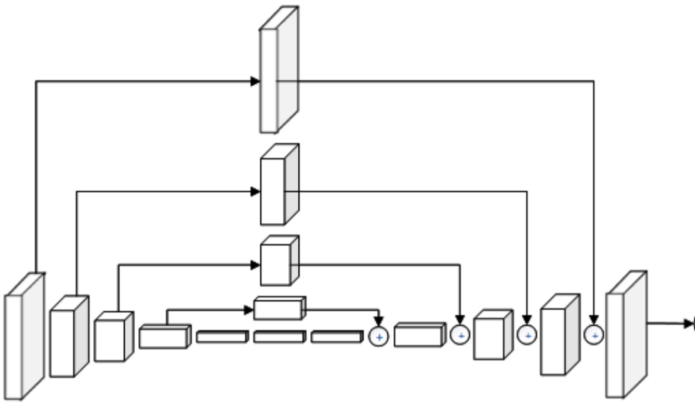


Figure 3.6: Hourglass architecture (Pavlakos et al., 2017).

final convolution layer. The elements discussed above make up the main body, which processes the inputs and produces output maps of the same size (Barad, 2020).

THe HRNet network used in this thesis consists of a first stage with four residual bottlenecks inspired by the ResNet architecture (He et al., 2016), followed by a second stage with one block, a third stage with four blocks and a fourth stage with three blocks. The network uses 32, 64, 128 and 256 channels from highest to lowest resolution parallel sub-networks, resulting in 28.5 M network parameters.

### HOURGLASS

The Hourglass network uses *pooling layers* to down sample and *skip-connections* between up and down sampling to aggregate multi-scale feature maps. Typically, one or more hourglass modules are stacked together in order to improve the heatmaps resolution and boost performance. Compared to the network proposed by Pavlakos et al. (2017), the hourglass architecture adopted in this thesis is composed of only one encoder/decoder block, constituting a single-stack hourglass module. This was chosen in order to reduce the network size and comply with the limitations in computing power which characterizes space-grade processors. The encoder includes six blocks, each including a convolutional layer formed by a fixed number of filter kernels of size $3\times3$, a batch normalization module and max pooling layer, whereas the six decoder blocks accomodate an up-sampling block in spite of max pooling. In the encoder stage, the initial image resolution is decreased by a factor of two, with this downsampling process continuing until reaching the lowest resolution of $4 \times 4$ pixels. An upsampling process follows in the decoder with each layer increasing the resolution by a factor of two and returning output heatmaps at the same resolution as the input image. Overall, the size of the 2D input image and the number of kernels per convolutional layer drive the total number of parameters. In the current analysis, an input size of $256 \times 256$ pixels is chosen, and 128 kernels are considered per convolutional layer, leading to a total of $\sim 1,800,000$ M trainable parameters. Compared

Table 3.1: Parameters of the camera used to generate the synthetic images in Cinema 4D©.

| Parameter | Value | Unit |
|---|---|---|
| Image resolution | 512×512 | pixels |
| Focal length | 3.9 | mm |
| Pixel size | $1.1 \cdot 10^{-5}$ | m |



to both the HRNet and a traditional stacked hourglass, this represents a reduction of more than an order of magnitude in network size.

### 3.3.2. Loss Function

During training, both the Single-stack Hourglass and the HRNet use the target heatmaps and the predicted heatmaps to compute the loss. For the heatmap representation, a simple Mean Squared Error (MSE) is generally used (Sun et al., 2019; Szegedy et al., 2016) for per-pixel confidence values between the predicted and the target heatmap. Given $k$ target heatmaps and the corresponding heatmaps predicted by the network, the loss function is given as:

$$L = \frac{1}{k}\sum_{h=1}^{k} L_h \quad ; \quad L_h = \frac{1}{m}\frac{1}{n}\sum_{i,j}\left(c_{i,j} - c'_{i,j}\right)^2, \tag{3.1}$$

where, $c_{i,j}$ is the target confidence, and $c'_{i,j}$ is the predicted confidence for a pixel at location $i, j$ in the heatmap of size $m \times n$.

### 3.3.3. Training and Evaluation

The proposed networks are optimized to locate 16 features of the Envisat spacecraft, consisting of the corners of the main body, the Synthetic-Aperture Radar (SAR) antenna, and the solar panel, respectively. Figure 3.7 illustrates the selected features for a specific target pose. From an evaluation perspective, a CNN could be simply trained on a training dataset and evaluated on a test dataset at the end of the last training *epoch*, in order to assess its detection performance on previously unseen images of the target spacecraft. In this context, multiple epochs are chosen in order to reiterate on the same images while improving on the detection robustness, with each epoch ending when the network has been trained on every image of the training dataset. However, an additional validation dataset is typically used during training to compute additional validation losses and avoid *overfitting*. In essence, overfitting would occur when the network returns highly accurate detections on the training dataset but performs poorly on any other dataset. The inclusion of an additional dataset during training would guarantee that the detection performance of the network can be monitored at the end of each epoch, in order to stop

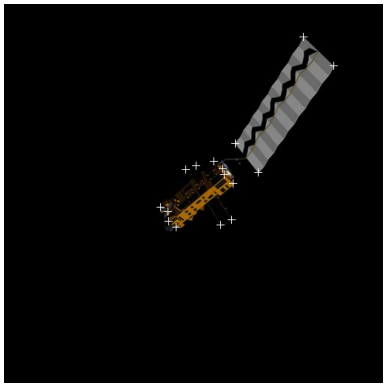


Figure 3.7: Illustration of the selected features for a given Envisat pose.



the training as soon as the validation losses start to increase.

For the training, validation, and test datasets, synthetic images of the Envisat spacecraft are rendered in the Cinema 4D© software. Table 4.2 lists the main camera parameters adopted. The Sun elevation and azimuth angles with respect to the camera frame are fixed at 30 degrees. Relative distances between camera and target are discretized every 30 m in the interval 90 m - 180 m, with the Envisat always located in the camera boresight direction in order to prevent some of the Envisat features from falling outside the camera field of view. Although being a conservative assumption, this allows to test the CNN performance in ideal servicer-target geometries during a rectilinear approach. Subsequently, relative attitudes are generated by discretizing the yaw, pitch, and roll angles of the target with respect to the camera by 10 degrees each. Together, these two choices were made in order to recreate several relative attitudes between the servicer and the target. The resulting database is then shuffled to randomize the images, and is ultimately split into training (18,000 images), validation (6,000 images), and test (6,000 images) datasets. Figure 3.8 shows a subset of the camera pose distribution for 100 representative training images, whereas Figure 3.9 illustrates some of the images included in the training dataset.

The Adam optimizer (Kingma and Ba, 2015) is used with a cosine decaying learning rate with initial value of $10^{-3}$ and decaying factor of 0.1. Finally, the network performance after training is assessed on the test dataset. A Tesla P100-PCIE-16GB Graphics Processing Unit (GPU) is used for both training and testing.

### 3.3.4. KEYPOINT DETECTION PERFORMANCE

Keypoint detection results are firstly reported for the Hourglass network on the Envisat test dataset. The performance is assessed in terms of Root Mean Squared Error (RMSE) between the ground truth (GT) and the $x, y$ coordinates of the extracted features, which is computed as:

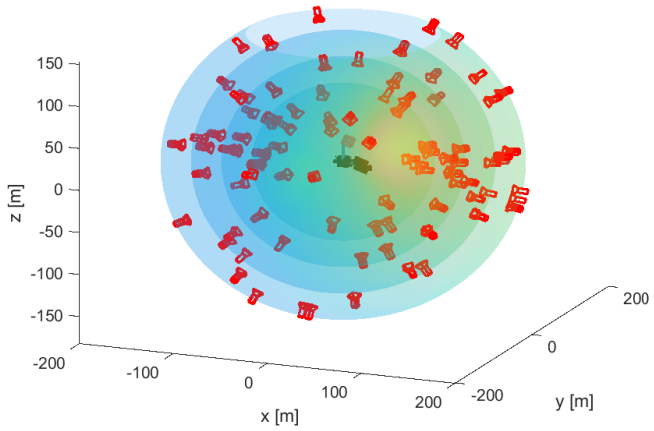



Figure 3.8: Illustration of the pose space discretization in the training dataset. The concentric spheres represent the discretization of the relative distance in the range 90 m - 180 m. Only 100 random relative camera poses are shown for clarity.

$$E_{\text{RMSE}} = \sqrt{\frac{\sum_{i=1}^{n_{tot}} \left[ (x_{GT,i} - x_i)^2 + (y_{GT,i} - y_i)^2 \right]}{n_{tot}}}. \tag{3.2}$$

Figure 3.10 illustrates a sample of four detection scenarios. In the upper-left figure, a good features detection characterized by a total RMSE of 1.37 pixels is presented. As can be seen, each detected feature deviates from its ground truth by less than 2 pixels. The upper-right figure also present a highly accurate detection scenario, in which the total RMSE is at subpixel level. Instead, in the lower-left image a scenario is presented in which the RMSE relates to a moderately accurate detection. Finally, the lower-right image shows a scenario in which the network is unable to correctly detect the last feature and returns a large RMSE. Specifically, the CNN detects a corner in the spacecraft antenna rather than one of the corners of the solar panel. This suggests that large RMSE might be associated to the incorrect detection of just one (or a few) keypoints.

In general, one key advantage of CNNs for feature detection can be identified in their capability to learn the relative position between features under a variety of illumination conditions and poses present in the training. As a result, both features which are not visible due to adverse illumination and features occulted by other parts of the target can be detected. Figure 3.11 shows the RMSE error over the test dataset for both the single-stack hourglass and the HRNet, whereas Table 3.2 reports the mean $\mu$ and standard deviation $\sigma$ of the associated histograms. As expected, the added complexity of HRNet translates into a more accurate detection of the selected features, thanks to the higher

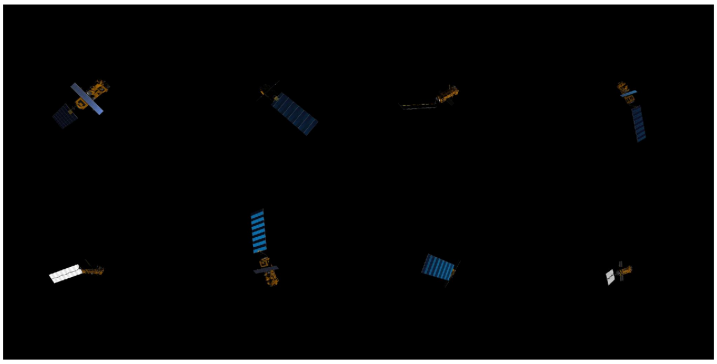


Figure 3.9: A montage of eight synthetic images selected from the training set.

number of parameters: only 4% of the test images are characterized by a RMSE above 5 pixels, as opposed to the 15% in the Single-stack Hourglass case.

Table 3.2: Mean $\mu$ and Standard Deviation $\sigma$ of the adopted networks over the Envisat test dataset.

| Network | No. Params | $\mu$ [pxl] | $\sigma$ [pxl] |
|---|---|---|---|
| Single-stack Hourglass | 1.8 M | 3.4 | 4.3 |
| HRNet | 28.5 M | 2.4 | 1.4 |

Although HRNet proves to return more accurate features, it is also believed that the larger RMSE scenarios returned by the Single-stack Hourglass can be properly handled, if a larger uncertainty can be associated to their corresponding heatmaps. As an example, a large RMSE could be associated to the inaccurate detection of only a few features which, if properly weighted, would not have a severe impact on the pose estimation step. This task can be performed by deriving a covariance matrix for each detected feature, in order to represent its detection uncertainty. Above all, this may prevent the pose solver and the navigation filter from trusting wrong detections by relying more on other accurate features. In this way, the navigation filter could handle inaccurate heatmaps while at the same time relying on a computationally-low CNN.

## 3.4. COVARIANCE COMPUTATION

Compared to the methods discussed in Section 3.1 (Cui et al., 2019; Ferraz et al., 2014a; Harvard et al., 2020), the proposed method derives a covariance matrix associated to each feature directly from the heatmaps detected by the CNN, rather than from the computation of the image gradient around each feature. In order to do so, the first step is to obtain a statistical population around the heatmap's peak. This is done by thresholding each heatmap image so that only the $x$- and $y$- location of heatmap's pixels are extracted.

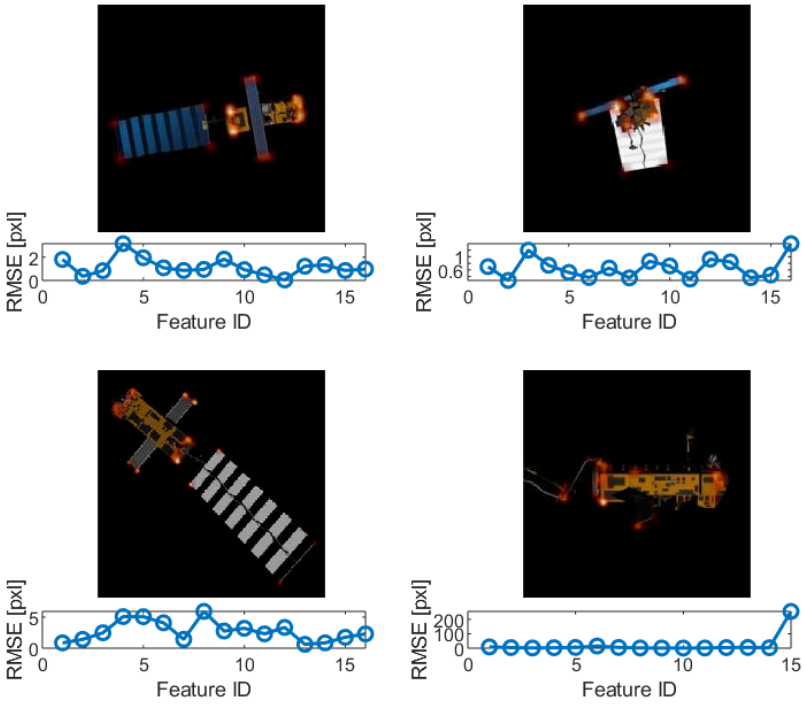



Figure 3.10: Four different test scenarios for feature detection.

Secondly, each pixel within the population is given a normalized weight $w_i$ based on the gray intensity $I_i$ at its location,

$$w_i = w_R R_i + w_G G_i + w_B B_i, \tag{3.3}$$

where $R, G, B$ are the components of the coloured image and $w_R, w_G, w_B$ are the weights assigned to each channel in order to obtain the grayscale intensity. This is done in order to give more weight to pixels which are particularly bright and close to the peak, and less weight to pixels which are very faint and far from the peak. Finally, the obtained statistical population of each feature is used to compute the weighted covariance between $x, y$ and consequently the covariance matrix $\boldsymbol{C}_i$,

$$\boldsymbol{C}_i = \begin{bmatrix} \text{cov}(x,x) & \text{cov}(x,y) \\ \text{cov}(y,x) & \text{cov}(y,y) \end{bmatrix}, \tag{3.4}$$

where

$$\text{cov}(x,y) = \sum_{i=1}^{n} w_i (x_i - p_x) \cdot (y_i - p_y) \tag{3.5}$$

and $n$ is the number of pixels in each feature's heatmap. In this work, the mean is replaced by the peak location $\boldsymbol{p} = (p_x, p_y)$ in order to represent a distribution around the peak of

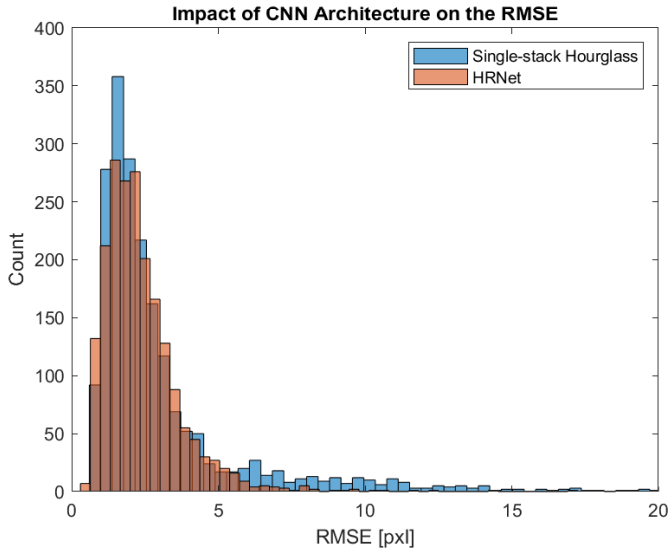



Figure 3.11: RMSE over the test dataset for the HRNet and the Single-stack Hourglass.

the detected feature, rather than around the heatmap's mean. This is particularly relevant when the heatmaps are asymmetric and their mean does not coincide with their peak.

Figure 5.19 shows the overall flow to obtain the covariance matrix for three different heatmap shapes. The ellipse associated to each features covariance is obtained by computing the eigenvalues $\lambda_x$ and $\lambda_y$ of the covariance matrix,

$$\left(\frac{x}{\lambda_x}\right)^2 + \left(\frac{y}{\lambda_y}\right)^2 = s, \tag{3.6}$$

where $s$ defines the scale of the ellipse and is derived from the confidence interval of interest, e.g. $s = 2.2173$ for a 68% confidence interval. As can be seen, different heatmaps can result in very different covariance matrices. Above all, the computed covariance can capture the different CNN uncertainty over $x, y$. Notice that, due to its symmetric nature, the covariance matrix can only represent bivariate normal distributions. As a result, asymmetrical heatmaps such as the one in the third scenario are approximated by Gaussian distributions characterized by an ellipse which might overestimate the heatmap's dispersion over some directions.

## 3.5. POSE ESTIMATION

The promising pose estimation results achieved in ADR scenarios in recent studies (Chen et al., 2019; Kisantal et al., 2020; Sharma and D'Amico, 2015, 2017) suggest that the EP$n$P method followed by Gauss-Newton refinement (Lepetit et al., 2009) is a viable method to estimate the pose from a set of detected features. This method solves the P$n$P problem in Equation 2.3 in closed-form with the EP$n$P algorithm, and uses the estimated pose

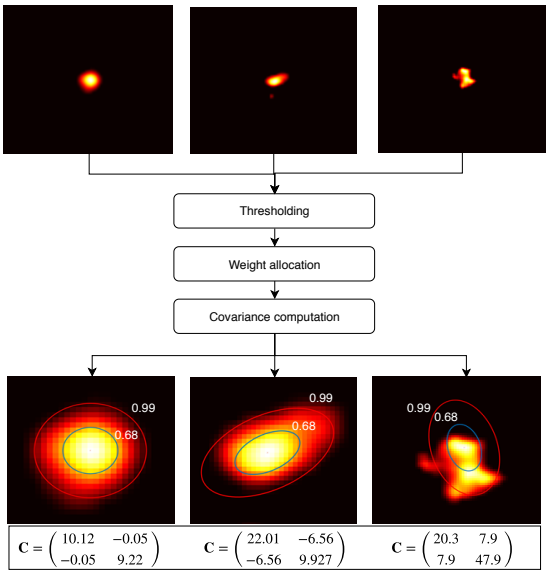



Figure 3.12: Schematic of the procedure followed to derive covariance matrices from CNN heatmaps. The displayed ellipses are derived from the computed covariances by assuming the confidence intervals $1\sigma = 0.68$ and $3\sigma = 0.99$.

as initial guess for an iterative pose refinement. The fundamental equation of the EP$n$P algorithm consists of rewriting the P$n$P problem by expressing the 3D model points as a weighted sum of four non-coplanar virtual control points $c_j$ in the camera frame C,

$$\boldsymbol{r}_i^C = \sum_{j=1}^{4} \alpha_{ij} \boldsymbol{c}_j^c, \tag{3.7}$$

where $\alpha_{ij}$ are the homogeneous barycentric coordinates computed from the 3D model points $\boldsymbol{r}_i^B$ through matrix inversion by assuming arbitrary coordinates of the control points $\boldsymbol{c}_j^B$ expressed in the body frame B (Lepetit et al., 2009). Using Equation 3.7, Equation 2.3 can be re-written in terms of the 12 unknown control point coordinates in the camera frame $[\hat{\alpha}_j^C, \hat{\beta}_j^C, \hat{\gamma}_j^C]^T$:

$$\omega_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} & K & \end{bmatrix} \sum_{j=1}^{4} \alpha_{ij} \begin{bmatrix} \hat{\alpha}_j^C \\ \hat{\beta}_j^C \\ \hat{\gamma}_j^C \end{bmatrix}. \tag{3.8}$$

By substituting the values of $\omega_i$ from the third row into the first two rows of Equation 3.8, two linear equations are formed for each corresponding pair of a 3D model point and image point,

$$\sum_{j=1}^{4} \alpha_{ij} f_x \hat{\alpha}_j - \alpha_{ij} u_i \hat{\gamma}_j = 0, \tag{3.9}$$

$$\sum_{j=1}^{4} \alpha_{ij} f_y \hat{\beta}_j - \alpha_{ij} v_i \hat{\gamma}_j = 0. \tag{3.10}$$

Eqs. 3.9-3.10 can then be written in compact form as

$$\boldsymbol{M}\boldsymbol{y} = \boldsymbol{0}, \tag{3.11}$$

where $\boldsymbol{M}$ is a known $2n \times 12$ matrix. It can be proven that the pose solution belongs to the kernel of $\boldsymbol{M}$, and therefore that it can be expressed as a linear combination of the columns of the right-singular vectors of $\boldsymbol{M}$ corresponding to the null singular values of $\boldsymbol{M}$ (Lepetit et al., 2009). As a result, an iterative refinement based on the Gauss-Newton method can be performed with little additional computational cost whilst improving the pose estimate. Note that the EP$n$P algorithm cannot return an estimate if less than four features are provided as input. As such, no pose estimate can be expected when a large amount of detected keypoints falls below the set threshold for the detection accuracy.

Unfortunately, the EP$n$P method does not have the capability to include feature uncertainties. As such, the heatmaps-based covariance derived in Section 3.4 cannot be directly used in the estimation process. To cope with this limitation, the CEPP$n$P method (Ferraz et al., 2014a) has been proposed to exploit statistical information by including feature covariances in the pose estimation. The first step of the CEPP$n$P algorithm is to represent the likelihood of each observed feature location $\boldsymbol{u}_i$ as

$$P(\boldsymbol{u}_i) = k \cdot e^{-\frac{1}{2}\Delta\boldsymbol{u}_i^T \boldsymbol{C}_{\boldsymbol{u}_i}^{-1} \Delta\boldsymbol{u}_i}, \tag{3.12}$$

where $\Delta\boldsymbol{u}_i$ is a small, independent and unbiased noise with expectation $E[\Delta\boldsymbol{u}_i] = \boldsymbol{0}$ and covariance $E[\Delta\boldsymbol{u}_i \Delta\boldsymbol{u}_i^T] = \sigma^2 \boldsymbol{C}_{\boldsymbol{u}_i}$ and $k$ is a normalization constant. Here, $\sigma^2$ represents the global uncertainty in the image, whereas $\boldsymbol{C}_{\boldsymbol{u}_i}$ is the 2x2 un-normalized covariance matrix representing the Gaussian distribution of the $i$th feature, computed from the CNN heatmaps. After some calculations (Ferraz et al., 2014a), the EP$n$P formulation in Equation 3.11 can be rewritten as

$$(\boldsymbol{N} - \boldsymbol{L})\boldsymbol{y} = \lambda \boldsymbol{y}. \tag{3.13}$$

This is an eigenvalue problem in which both $\boldsymbol{N}$ and $\boldsymbol{L}$ matrices are a function of $\boldsymbol{y}$ and $\boldsymbol{C}_{\boldsymbol{u}_i}$. The problem is solved iteratively by means of the closed-loop EPP$n$P solution (Ferraz et al., 2014b) for the four control points, assuming no feature uncertainty. Once $\boldsymbol{y}$ is estimated, the pose is computed by solving the generalized Orthogonal Procrustes problem used in the EPP$n$P (Ferraz et al., 2014b).

## 3.6. Navigation Filter

In the proposed navigation system, the so-called Multiplicative Extended Kalman Filter (MEKF) is used. Remarkably, other works (Harvard et al., 2020; Pasqualetto Cassinis et al., 2020) adopted a standard formulation of the EKF that propagates the pose, expressed in terms of relative position and quaternions, as well as the relative translational and rotational velocities (prediction step), correcting the prediction with the measurements obtained from the monocular camera (correction step). However, the quaternion set consists of four parameters to describe the 3DOF attitude, hence one of its parameters is deterministic. As reported by Tweddle and Saenz-Otero (2015) and Sharma and D'Amico (2017), this makes the covariance matrix of a quaternion have one eigenvalue that is exactly zero. As a result, the entire state covariance propagated by the filter may become non-positive-definite and lead to the divergence of the filter. The MEKF, introduced for the first time by Lefferts et al. (1982), aims at solving the above issue by using two different parametrizations of the relative attitude. A three elements error parametrization, expressed in terms of quaternions, is propagated and corrected inside the filter to return an estimate of the attitude error. At each estimation step, this error estimate is used to update a reference quaternion and is reset to zero for the next iteration. Notably, the reset step prevents the attitude error parametrization from reaching singularities, which generally occur for large angles.

### 3.6.1. Propagation Step

A standard EKF state vector for pose estimation is composed of the pose between the servicer and the target, as well as the relative translational and rotational velocities $\boldsymbol{v}$ and $\boldsymbol{\omega}$. Under the assumption that the camera frame onboard the servicer is co-moving with the LVLH frame, with the camera boresight aligned with the along-track direction, this translates into

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{t}^C & \boldsymbol{v} & \boldsymbol{q} & \boldsymbol{\omega} \end{bmatrix}^T, \tag{3.14}$$

where $\boldsymbol{q} = [q_0 \quad \boldsymbol{q_v}]$ is the quaternion set that represents the relative attitude. Notice that the assumption of the camera co-moving with the LVLH is made only to focus on the navigation aspects rather than on the attitude control of the servicer. Therefore, the application of the filter can be extended to other scenarios, if attitude control is included in the system.

In the MEKF, the modified state vector propagated inside the filter becomes

$$\tilde{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{t}^C & \boldsymbol{v} & \boldsymbol{a} & \boldsymbol{\omega} \end{bmatrix}^T, \tag{3.15}$$

where $\boldsymbol{a}$ is four times the Modified Rodrigues Parameters (MRP) error $\delta\boldsymbol{\sigma}$,

$$\boldsymbol{a} = 4\delta\boldsymbol{\sigma} = 4\frac{\delta\boldsymbol{q_v}}{1+\delta q_0}. \tag{3.16}$$

in which $\delta\boldsymbol{q}$ represents the quaternion error. The discrete attitude propagation step is derived by linearizing $\dot{\boldsymbol{a}}$ around $\boldsymbol{a} = \boldsymbol{0}_{3\times1}$ and assuming small angle rotations( Tweddle and Saenz-Otero, 2015),

$$\dot{\boldsymbol{a}} = \frac{1}{2}[\boldsymbol{\omega}\times]\boldsymbol{a} + \boldsymbol{\omega}. \tag{3.17}$$

As a result, the discrete linearized propagation of the full state becomes

$$\tilde{\boldsymbol{x}}_k = \boldsymbol{\Phi}_k \tilde{\boldsymbol{x}}_{k-1} + \boldsymbol{\Gamma}_k \boldsymbol{Q}_k, \tag{3.18}$$

where $\boldsymbol{Q}_k$ represents the process noise and (Tweddle and Saenz-Otero, 2015)

$$\boldsymbol{\Phi}_k = \begin{bmatrix} \boldsymbol{\Phi}_{CW} & \boldsymbol{0}_{6\times6} \\ \boldsymbol{0}_{6\times6} & \boldsymbol{\Phi}_{a,\omega} \end{bmatrix}, \tag{3.19}$$

where

$$\boldsymbol{\Phi}_{CW} = \begin{bmatrix} \boldsymbol{\Phi}_{rr} & \boldsymbol{\Phi}_{rv} \\ \boldsymbol{\Phi}_{vr} & \boldsymbol{\Phi}_{vv} \end{bmatrix}, \tag{3.20}$$

$$\boldsymbol{\Phi}_{rr} = \begin{bmatrix} 1 & 0 & 6(\Delta\theta - \sin\Delta\theta) \\ 0 & \cos\Delta\theta & 0 \\ 0 & 0 & 4 - 3\cos\Delta\theta \end{bmatrix}, \tag{3.21}$$

$$\boldsymbol{\Phi}_{rv} = 1/\omega_s \begin{bmatrix} 4\sin\Delta\theta - 3\Delta\theta & 0 & 2(1-\cos\Delta\theta) \\ 0 & \sin\Delta\theta & 0 \\ 2(\cos\Delta\theta - 1) & 0 & \sin\Delta\theta \end{bmatrix}, \tag{3.22}$$

$$\boldsymbol{\Phi}_{vr} = \begin{bmatrix} 0 & 0 & 6\omega_s(1-\cos\Delta\theta) \\ 0 & \omega_s\sin\Delta\theta & 0 \\ 0 & 0 & 3\omega_s\sin\Delta\theta \end{bmatrix}, \tag{3.23}$$

$$\boldsymbol{\Phi}_{vv} = \begin{bmatrix} 4\cos\Delta\theta - 3 & 0 & 2\sin\Delta\theta \\ 0 & \cos\Delta\theta & 0 \\ -2\sin\Delta\theta & 0 & \cos\Delta\theta \end{bmatrix}, \tag{3.24}$$

$$\boldsymbol{\Phi}_{a,\omega} = \begin{bmatrix} \boldsymbol{e}^{\frac{1}{2}[\omega\times]\Delta t} & \int_0^{\Delta t} \boldsymbol{e}^{\frac{1}{2}[\omega\times]\tau} d\tau \\ \boldsymbol{0}_{3\times3} & \boldsymbol{I}_{3\times3} \end{bmatrix}, \tag{3.25}$$

$$\boldsymbol{\Gamma}_{k+1} = \begin{bmatrix} \frac{1}{2m}\boldsymbol{I}_{3\times3}\Delta t^2 & \boldsymbol{0}_{3\times3} \\ \frac{1}{m}\boldsymbol{I}_{3\times3}\Delta t & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \Delta t\int_0^{\Delta t} \boldsymbol{e}^{\frac{1}{2}[\omega\times]\tau}\boldsymbol{J}^{-1} d\tau \\ \boldsymbol{0}_{3\times3} & \boldsymbol{J}^{-1}\Delta t \end{bmatrix}. \tag{3.26}$$

The terms $\omega_s$ and $\Delta\theta$ in Equation 3.20 represent the servicer argument of perigee and true anomaly variation from time $t_0$ to $t$, respectively, the terms $m$ and $\boldsymbol{J}$ in Equation 3.26 are the mass and the inertia matrix of the target spacecraft, and $\Delta t$ is the propagation step. In Tweddle and Saenz-Otero (2015), the integral terms in Eqs. 3.25-3.26 are solved by creating a temporary linear system from Equation 3.17, augmented with the angular velocity and the process noise. The State Transition Matrix of this system is then solved numerically with the matrix exponential.

### 3.6.2. Correction Step

At this stage, the propagated state $\tilde{\boldsymbol{x}}_k$ is corrected with the measurements $\boldsymbol{z}$ to return an update of the state $\hat{\boldsymbol{x}}_k$. In a loosely-coupled filter, these measurements are represented by the pose between the servicer and the target spacecraft, obtained by solving the P$n$P problem with the CEPP$n$P solver described in Section 3.5. In this case, a pseudomeasurements vector is derived by transforming the relative quaternion set into the desired attitude error $\boldsymbol{a}$,

$$\delta \boldsymbol{q}_z = \boldsymbol{q}_z \otimes \boldsymbol{q}_{\text{ref}_k^*} \rightarrow \boldsymbol{a} = 4\frac{\delta \boldsymbol{q}_v}{1+\delta q_0}, \tag{3.27}$$

$$\boldsymbol{z}_k = \begin{bmatrix} \boldsymbol{t}^C \\ \boldsymbol{a} \end{bmatrix} = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{V}_k = \begin{bmatrix} \boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times3} \end{bmatrix} \begin{bmatrix} \boldsymbol{t}^C \\ \boldsymbol{v} \\ \boldsymbol{a} \\ \boldsymbol{\omega} \end{bmatrix}_k + \begin{bmatrix} \boldsymbol{V}_r \\ \boldsymbol{V}_a \end{bmatrix}_k. \tag{3.28}$$

In Equation 3.27, $\otimes$ denotes the quaternion product. Conversely, in a tightly-coupled filter the measurements are represented by the pixel coordinates of the detected features,

$$\boldsymbol{z} = \begin{bmatrix} x_1, y_1 & \dots & x_n, y_n \end{bmatrix}^T. \tag{3.29}$$

Referring to Eqs. 2.1-2.2, this translates into the following equations for each detected point $\boldsymbol{p}_i$:

$$\boldsymbol{h}_i = \begin{bmatrix} \frac{x_i^C}{z_i^C} f_x + C_x, \frac{y_i^C}{z_i^C} f_y + C_y \end{bmatrix}^T, \tag{3.30}$$

$$\mathbf{r}^C = \boldsymbol{q} \otimes \boldsymbol{r}_i^B \otimes \boldsymbol{q}^* + \mathbf{t}^C, \tag{3.31}$$

where $\boldsymbol{q}^*$ is the quaternion conjugate. As a result, the measurement update equation can be written as

$$\boldsymbol{z}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{V} = \begin{bmatrix} \boldsymbol{H}_{\boldsymbol{t}^C,i} & \boldsymbol{0}_{2n\times3} & \boldsymbol{H}_{\boldsymbol{a},i} & \boldsymbol{0}_{2n\times3} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{H}_{\boldsymbol{t}^C,n} & \boldsymbol{0}_{2n\times3} & \boldsymbol{H}_{\boldsymbol{a},n} & \boldsymbol{0}_{2n\times3} \end{bmatrix} \begin{bmatrix} \boldsymbol{t}^C \\ \boldsymbol{v} \\ \boldsymbol{a} \\ \boldsymbol{w} \end{bmatrix}_k + \begin{bmatrix} \boldsymbol{V}_r \\ \boldsymbol{V}_a \end{bmatrix} \tag{3.32}$$

and the Jacobian $\boldsymbol{H}_k$ of the observation model with respect of the state vector is a $2n\times13$ matrix whose elements are

$$\boldsymbol{H}_{\boldsymbol{t}^C,i} = \boldsymbol{H}_i^{\text{int}} \cdot \boldsymbol{H}_{\boldsymbol{t}^C,i}^{\text{ext}} \tag{3.33}$$

$$\boldsymbol{H}_{\boldsymbol{a},i} = \boldsymbol{H}_i^{\text{int}} \cdot \boldsymbol{H}_{\boldsymbol{a},i}^{\text{ext}} = \boldsymbol{H}_i^{\text{int}} \cdot \boldsymbol{H}_{\boldsymbol{q},i}^{\text{ext}} \cdot \boldsymbol{H}_{\boldsymbol{a}}^{\boldsymbol{q}} \tag{3.34}$$

$$\boldsymbol{H}_i^{\text{int}} = \frac{\partial \boldsymbol{h}_i}{\partial \boldsymbol{r}_i^C} = \begin{bmatrix} \frac{f_x}{z_i^C} & 0 & -\frac{f_x}{(z_i^C)^2} x_i^C \\ 0 & \frac{f_y}{z_i^C} & -\frac{f_y}{(z_i^C)^2} y_i^C \end{bmatrix} \tag{3.35}$$

$$\boldsymbol{H}^{\text{ext}}_{\boldsymbol{q},i} = \frac{\partial \boldsymbol{r}^C_i}{\partial \boldsymbol{q}} = \frac{\partial\left(\boldsymbol{q} \otimes \boldsymbol{r}^B_i \otimes \boldsymbol{q}^*\right)}{\partial \boldsymbol{q}}; \quad \boldsymbol{H}^{\text{ext}}_{\boldsymbol{t}^C,i} = \frac{\partial \boldsymbol{r}^C_i}{\partial \boldsymbol{t}^C} = \boldsymbol{I}_3 \tag{3.36}$$

$$\boldsymbol{H}^{\boldsymbol{q}}_{\boldsymbol{a}} = \frac{\partial\left(\delta\boldsymbol{q} \otimes \boldsymbol{q}_{\text{ref}}\right)}{\partial \boldsymbol{a}} = \frac{\partial\left(\boldsymbol{Q}_{\text{ref}}\delta\boldsymbol{q}\right)}{\partial \boldsymbol{a}} = \boldsymbol{Q}_{\text{ref}}\frac{\partial\left(\delta\boldsymbol{q}\right)}{\partial \boldsymbol{a}} \tag{3.37}$$

$$\boldsymbol{Q}_{\text{ref}} = \begin{bmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & q_3 & -q_2 \\ q_2 & -q_3 & q_0 & q_1 \\ q_3 & q_2 & -q_1 & q_0 \end{bmatrix}_{\text{ref}}. \tag{3.38}$$

The partial derivatives of the differential quaternion set with respect to the attitude error are computed from the relation between the attitude error $\boldsymbol{a}$ and the differential quaternion set $\delta\boldsymbol{q}$,

$$\delta q_0 = \frac{16 - \|\boldsymbol{a}\|^2}{16 + \|\boldsymbol{a}\|^2} \quad \delta\boldsymbol{q}_{\boldsymbol{v}} = 8\frac{\boldsymbol{a}}{16 + \|\boldsymbol{a}\|^2} \tag{3.39}$$

$$\frac{\partial\left(\delta\boldsymbol{q}\right)}{\partial \boldsymbol{a}} = \frac{8}{\left(16 + \|\boldsymbol{a}\|^2\right)^2}\begin{bmatrix} -8a_1 & -8a_2 & -8a_3 \\ 16 + \|\boldsymbol{a}\|^2 - 2a_1^2 & -2a_1a_2 & -2a_1a_3 \\ -2a_2a_1 & 16 + \|\boldsymbol{a}\|^2 - 2a_2^2 & -2a_2a_3 \\ -2a_3a_1 & -2a_3a_2 & 16 + \|\boldsymbol{a}\|^2 - 2a_3^2 \end{bmatrix}. \tag{3.40}$$

In the tightly-coupled filter, the measurement error covariance matrix $\boldsymbol{R}_k$ is a time-varying block diagonal matrix constructed with the heatmaps-derived covariances $\boldsymbol{C}_i$ in Equation 5.11,

$$\boldsymbol{R}_k = \begin{bmatrix} \boldsymbol{C}_1 & & \\ & \ddots & \\ & & \boldsymbol{C}_n \end{bmatrix}. \tag{3.41}$$

Notice that $\boldsymbol{C}_i$ can differ for each feature in a given frame as well as vary over time. Such heatmaps-derived covariance matrix can capture the statistical distribution of the measured features and improve the measurement update step of the navigation filter. Conversely, in the loosely-coupled filter $\boldsymbol{R}$ represents the uncertainty in the pose estimation step and hence it is not directly related to the CNN heatmaps. A constant value is therefore chosen based on the pose estimation accuracy observed for the test dataset.

Finally, the updated state estimate $\hat{\boldsymbol{x}}_k$ is obtained from the propagated state $\tilde{\boldsymbol{x}}_k$, the residuals $\tilde{\boldsymbol{y}}$, and the Kalman Gain $\boldsymbol{K}$,

$$\tilde{\boldsymbol{y}} = \boldsymbol{z} - \boldsymbol{h}(\tilde{\boldsymbol{x}}_k), \tag{3.42}$$

$$\boldsymbol{K} = \boldsymbol{P}_k\boldsymbol{H}_k^T\left(\boldsymbol{H}_k\boldsymbol{P}_k\boldsymbol{H}_k^T + \boldsymbol{R}_k\right)^{-1}, \tag{3.43}$$

$$\hat{\boldsymbol{x}}_k = \tilde{\boldsymbol{x}}_k + \boldsymbol{K}\tilde{\boldsymbol{y}}. \tag{3.44}$$

### 3.6.3. Reset Step

In the reset step, the reference quaternion $\boldsymbol{q}_{\text{ref}}$ is updated with the attitude error estimate $\hat{\boldsymbol{a}}_{\boldsymbol{p}}$ in order to obtain the estimated quaternion, and the new attitude error is set to zero,

$$\hat{\boldsymbol{q}}_k = \delta\boldsymbol{q}(\hat{\boldsymbol{a}}) \otimes \boldsymbol{q}_{\text{ref}_k}, \tag{3.45}$$

$$\hat{\boldsymbol{a}} = \mathbf{0}_{3\times 1}, \tag{3.46}$$

$$\boldsymbol{q}_{\text{ref}_{k+1}} = \hat{\boldsymbol{q}}_k. \tag{3.47}$$

The obtained estimated quaternion set $\hat{\boldsymbol{q}}_k$ is then compared to the real quaternion set to assess the angle accuracy of the filter.

## 3.7. Simulations

In this section, the simulation environment and the results are presented. Firstly, the impact of including a heatmaps-derived covariance in the pose estimation step is addressed by comparing the CEPP$n$P method with a standard solver which does not account for feature uncertainty. The weights in Equation 3.3 are selected based on the standard RGB-to-grayscale conversion ($w_R = 0.299$, $w_G = 0.587, w_B = 0.114$). Secondly, the performance of the MEKF is evaluated by comparing the convergence profiles with a heatmaps-derived covariance matrix against covariance matrices with constant covariances. Initialization is provided by the CEPP$n$P for all the scenarios.
Two separate error metrics are adopted in the evaluation, in accordance with Sharma and D'Amico, 2019. Firstly, the translational error between the estimated relative position $\hat{\boldsymbol{t}}^C$ and the ground truth $\boldsymbol{t}_C$ is computed as

$$E_T = |\boldsymbol{t}^C - \hat{\boldsymbol{t}}^C|. \tag{3.48}$$

A comparable metric is also applied for the translational and rotational velocities estimated in the navigation filter. Secondly, the attitude accuracy is measured in terms of the Euler axis-angle error between the estimated quaternion $\hat{\boldsymbol{q}}$ and the ground truth $\boldsymbol{q}$,

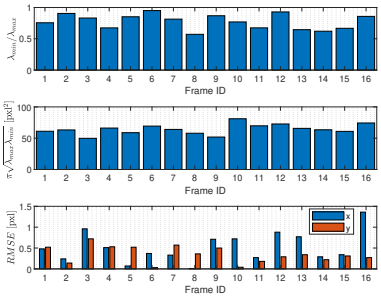
$$\boldsymbol{\beta} = \begin{bmatrix}\beta_s & \boldsymbol{\beta_v}\end{bmatrix} = \boldsymbol{q} \otimes \hat{\boldsymbol{q}}, \tag{3.49}$$

$$E_R = 2\arccos(|\beta_s|). \tag{3.50}$$

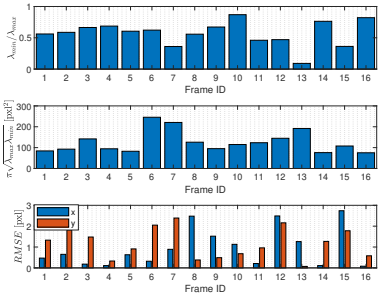
The Euler axis-angle error expresses the magnitude of the rotation that aligns the true target body frame B with the estimated target rotation.

### 3.7.1. Pose Estimation

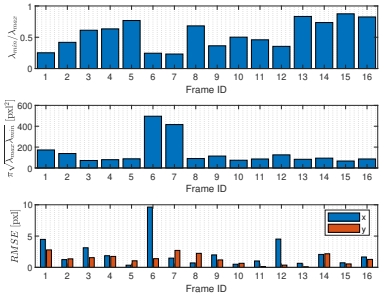
Three representative scenarios are selected from the CNN test dataset for a preliminary evaluation of the Single-stack Hourglass performance. These scenarios were chosen in order to analyze different heatmap's distributions around the detected features. A comparison is made between the proposed CEPP$n$P and the EPP$n$P. Figure 3.13 shows the characteristics of the covariance matrices derived from the predicted heatmaps. Here,




(a) Test scenario 1




(b) Test scenario 2




(c) Test scenario 3

Figure 3.13: Characteristics of the ellipses derived from the covariance matrices for the three selected scenarios.

the ratio between the minimum and maximum eigenvalues of the associated covariances is represented against the ellipse's area and the RMSE between the Ground Truth (GT) and the x, y coordinates of the extracted features,

$$E_{\text{RMSE},i} = \sqrt{(x_{GT,i} - x_i)^2 + (y_{GT,i} - y_i)^2}. \tag{3.51}$$

Notably, interesting relations can be established between the three quantities reported in the figure. In the first scenario, the correlation between the sub-pixel RMSE and the large eigenvalues ratio suggests that a very accurate CNN detection can be associated with circular-shaped heatmaps. Moreover, the relatively low ellipse's areas indicate that, in general, small heatmaps are expected for an accurate detection. Conversely, in the second scenario the larger ellipse's area correlates with a larger RMSE. Furthermore, it can be seen that the largest difference between the $x$- and $y$- components of the RMSE occurs either for the most eccentric heatmap (ID 13) or for the one with the largest area (ID 6). The same behaviour can be observed in the last scenario, where the largest RMSE coincides with a large, highly eccentric heatmap.

Table 3.3 lists the pose estimation results for the three scenarios. As anticipated in Figure 3.13, the statistical information derived from the heatmaps in the first scenario

Table 3.3: Single-stack Hourglass Pose Estimation performance results for the selected three representative scenarios.

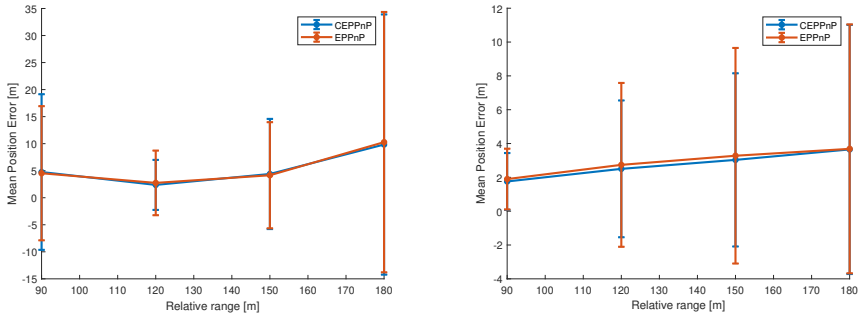
| Metric | Scenario | CEPP$n$P | EPP$n$P |
|---|---|---|---|
| $E_T$ [m] | 1 | [0.18 0.22 0.24] | [0.17 0.22 0.24] |
| | 2 | [0.35 0.41 0.59] | [0.14 0.4 22.8] |
| | 3 | [0.49 0.12 1.41] | [0.56 0.16 5.01] |
| $E_R$ [deg] | 1 | 0.36 | 0.35 |
| | 2 | 0.75 | 6.08 |
| | 3 | 1.99 | 2.72 |



is uniform for all the features, due to the very accurate CNN detection. As a result, the inclusion of features covariance in the CEPP$n$P solver does not help refining the estimated pose. Both solvers are characterized by the same pose accuracy.
Not surprisingly, the situation changes as soon as the heatmaps are not uniform across the feature IDs. Due to its capability of accomodating feature uncertainties in the estimation, the CEPP$n$P method outperforms the EPP$n$P for the remaining scenarios. In other words, the CEPP$n$P solver proves to be more robust against inaccurate CNN detections by accounting for a reliable representation of the features covariance.

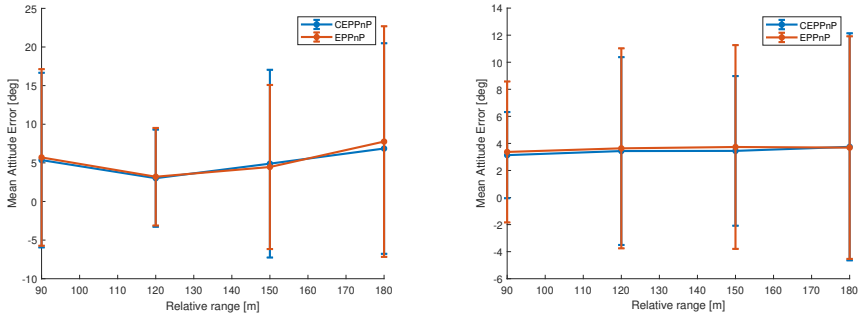
Next, the previous comparison is extended to the entire test dataset as well as to HRNet, by computing the mean and standard deviation of the estimated relative position and attitude as a function of the relative range, respectively. This is represented in Figs. 3.14-3.15. First of all, it can be seen that the pose accuracy of the CEPP$n$P solver in the Single-stack Hourglass scenario does not improve compared to the EPP$n$P, as opposed to the ideal behaviour reported in Table 3.3. There are two potential causes of this behaviour. On the one hand, most of the test images characterized by a large RMSE (Figure 3.11) could not return statistically-meaningful heatmaps that would help the CEPP$n$P solver. This could be due to multiple heatmaps or highly inaccurate detections in which two different corners are confused with each other. On the other hand, this could be a direct consequence of the large relative ranges considered in this work. As already reported by Park et al., 2019 and Sharma and D'Amico, 2017, a decreasing performance of EPP$n$P is indeed expected for increasing relative distances, due to the nonlinear relation between the pixel location of the detected features and $z^C$ in Equation 2.2. In other words, relatively large pixel errors could lead to inaccurate pose estimates for large relative distances, independently of the use of either CEPP$n$P or EPP$n$P.

Furthermore, it can be seen from a different comparison level that both the mean and standard deviation of the estimated pose are improved, when HRNet is used prior to the P$n$P solver (Figure 3.14b-3.15b). Again, this is a direct consequence of the smaller RMSE reported in Figure 3.11. As a result, the above-mentioned degradation of the pose estimation accuracy for increasing relative ranges is less critical for HRNet. Notice also that, despite an actual improvement of CEPP$n$P over EPP$n$P can be seen in the HRNet scenario, the improvements in both the mean and standard deviation of the estimation error are relatively small at large relative distances. This is considered to be related to the




(a) Single-stack Hourglass (b) HRNet

Figure 3.14: Pose Estimation Results - The standard deviation of the position error $E_T$ is depicted as the length of each error bar above and below the mean error $E_T$.




(a) Single-stack Hourglass (b) HRNet

Figure 3.15: Pose Estimation Results - The standard deviation of the attitude error $E_R$ is depicted as the length of each error bar above and below the mean error $E_R$.

fact that HRNet returns circular heatmaps for most of the detected features, due to its higher detection accuracy compared to the Single-stack Hourglass.

Notably, it is important to assess how well the pose estimation system can scale when tested on datasets different than the Envisat one. To this aim, the proposed heatmaps-based scheme was benchmarked on the SPEED dataset, in order to compare its pose accuracy against standard as well as CNN-based systems (Chen et al., 2019; Park et al., 2019; Sharma, Beierle, et al., 2018). The reader is referred to Barad, 2020, p. 115 for a comprehensive quantitative analysis of such comparison. The results demonstrated that the performance of the proposed pipeline, based on extracting feature heatmaps and using the CEPP$n$P solver, compares well with the state-of-the-art pose estimation systems.

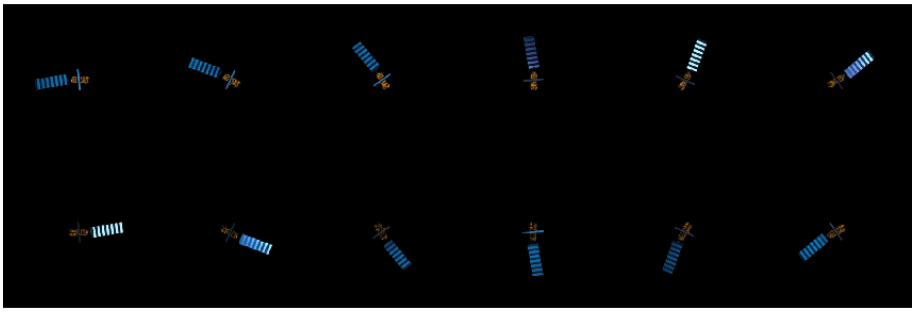


Figure 3.16: Montage of the selected vbar approach scenario. Images are shown every 6 s for clarity.

### 3.7.2. NAVIGATION FILTER

To assess the performance of the proposed MEKF, a rendezvous scenarios with Envisat is rendered in Cinema 4D©. This is a perturbation-free vbar trajectory characterized by a relative velocity $\|\boldsymbol{v}\| = 0$ m/s. The Envisat performs a roll rotation of $\|\boldsymbol{\omega}\| = 5$ deg/s, with the servicer camera frame aligned with the LVLH frame. Table 3.4 lists the initial conditions of the trajectory, whereas Figure 3.16 shows some of the associated rendered 2D images. It is assumed that the images are made available to the filter every 2 seconds for the measurement update step, with the propagation step running at 1 Hz. In both scenarios, the MEKF is initialized with the CEPP$n$P pose solution at time $t_0$. The other elements of the initial state vector are randomly chosen assuming a standard deviation of 1 mm/s and 1 deg/s for all the axes of terms $(\hat{\boldsymbol{v}}_0 - \boldsymbol{v})$ and $(\hat{\boldsymbol{\omega}}_0 - \boldsymbol{\omega})$, respectively. Table 3.5 reports the initial conditions of the filter.

Table 3.4: Vbar approach scenario. The attitude is represented in terms of ZYX Euler angles for clarity. Note that the camera boresight is the Y-axis of the LVLH frame.

| $\boldsymbol{\theta}_0$ [deg] | $\boldsymbol{\omega}_0$ [deg/s] | $\boldsymbol{t}_0^C$ [m] | $\boldsymbol{v}_0$ [mm/s] |
|---|---|---|---|
| [-180 30 -80]$^T$ | [-2.5 -4.3 0.75]$^T$ | [0 150 0]$^T$ | [0 0 0]$^T$ |

Table 3.5: Initial state vector in the MEKF. Here, HG refers to the Single-stack Hourglass architecture.

| **CNN** | $\hat{\boldsymbol{\theta}}_0$ [deg] | $\hat{\boldsymbol{\omega}}_0$ [deg/s] | $\hat{\boldsymbol{t}}_0^C$ [m] | $\hat{\boldsymbol{v}}_0$ [mm/s] |
|---|---|---|---|---|
| HG | [-180.2 28.7 -80.6]$^T$ | [-2.1 -4.1 0.1]$^T$ | [0.1 149.7 0.1]$^T$ | [2.8 -1.3 3]$^T$ |
| HRNet | [-179.5 33.5 -79.7]$^T$ | [-2.1 -4.1 0.1]$^T$ | [-0.1 149.8 -0.1]$^T$ | [2.8 -1.3 3]$^T$ |

Figs. 3.17-3.18 show the convergence profiles for the translational and rotational states in the tightly- and loosely-coupled MEKF, respectively. Moreover, a Monte Carlo simulation with 1,000 runs was performed to assess the robustness of the filter estimate against varying the initial state $\hat{\boldsymbol{x}}_0$. Table 3.6 lists the standard deviation chosen for the deviation from the true initial state of the filter. The distribution follows a Gaussian profile with true-state mean. For the attitude initial error, the initial reference quaternion $\boldsymbol{q}_{\text{ref}_0}$

Table 3.6: Standard deviation of Monte Carlo variables.

| $\sigma_{\Delta\phi_0}$ [deg] | $\sigma_{\boldsymbol{\omega}_0}$ [deg/s] | $\sigma_{\boldsymbol{t}_0^C}$ [m] | $\sigma_{\boldsymbol{v}_0}$ [mm/s] |
|---|---|---|---|
| 10 | 1 | $[1\ 10\ 1]^T$ | 10 |

is perturbed by introducing a random angular error around the correct Euler axis (Solà, 2017, p. 44),

$$\Delta\boldsymbol{\phi}_0 = \Delta\phi_0 \boldsymbol{q}_{\boldsymbol{v}} \tag{3.52}$$

$$\boldsymbol{q}_{\text{ref}_0} = \boldsymbol{q}_0 \otimes \begin{bmatrix} 1 \\ \frac{1}{2}\Delta\boldsymbol{\phi}_0 \end{bmatrix}. \tag{3.53}$$

Table 3.7 reports the mean of the steady-state pose estimates together with their standard deviation. From these results, important insights can be gained on two different levels of the comparison.

Table 3.7: Monte Carlo Simulation Results. The mean and standard deviation of the pose errors are taken from the absolute errors at filter steady state for each Monte Carlo run.

| **Single-stack Hourglass** | | | | |
|---|---|---|---|---|
| **MEKF** | $E_{T_1}$ [m] | $E_{T_2}$ [m] | $E_{T_3}$ [m] | $E_R$ [deg] |
| Tigthly-coupled | 0.1182 ± 6.3E-4 | 0.03 ± 0.0024 | 0.096 ± 5.4E-4 | 1.33 ± 0.03 |
| Loosely-coupled | 0.1 ± 2E-7 | 0.33 ± 3E-7 | 0.01 ± 1E-4 | 4.7 ± 12.6 |
| **HRNet** | | | | |
| **MEKF** | $E_{T_1}$ [m] | $E_{T_2}$ [m] | $E_{T_3}$ [m] | $E_R$ [deg] |
| Tigthly-coupled | 0.0683 ± 5.2E-4 | 0.03 ± 0.014 | 0.01 ± 3.4E-5 | 4.3 ± 0.03 |
| Loosely-coupled | 0.0075 ± 1E-4 | 0.36 ± 6.3E-5 | 0.002 ± 4E-4 | 0.93 ± 3.2E-7 |

On a CNN performance level, the results in Figure 3.17 show that a slightly worse cross-track estimate of the Single-stack Hourglass is compensated by a more accurate estimate of the relative attitude. Given the limited impact of these estimation errors at the relatively large inter-satellite range of 150 m, these results suggest that the Single-stack Hourglass has a comparable performance with the HRNet for the selected scenario. Next, on a filter architecture level, a comparison between Figs. 3.17-3.18 illustrate the different convergence pattern between the tightly- and loosely-coupled MEKF. Most importantly, it can be seen that the loosely-coupled estimate of the relative along-track position is characterized by a bias which is not present in the tightly-coupled estimate. This occurs due to the decoupling of the translational and rotational states, reflected in the Jacobian $\boldsymbol{H}_k$ in Equation 3.28. As a result, the relative position is estimated without accounting for the attitude measurements and viceversa. In other words, the creation of pseudomeasurements of the pose prior to the loosely-coupled filter leads to two separate translational and rotational estimates. Conversely, in the tightly-coupled filter the full statistical information is enclosed in the detected features, and can be used to simultaneously refine

both the translational and the rotational states. Moreover, a close inspection of the Single-stack Hourglass attitude estimates in Table 3.7 suggests that the tightly-coupled MEKF is characterized by a lower standard deviation, highlighting a better robustness with respect to the initial conditions of the filter when compared to the loosely-coupled MEKF. Note that, due to the higher accuracy of HRNet in feature detection - and hence also in pose estimation, this is not observed for the latter CNN. In conclusion, a tightly-coupled architecture is expected to return higher pose accuracies if simplified CNNs, such as the proposed single-stack hourglass, are implemented at a feature detection level.

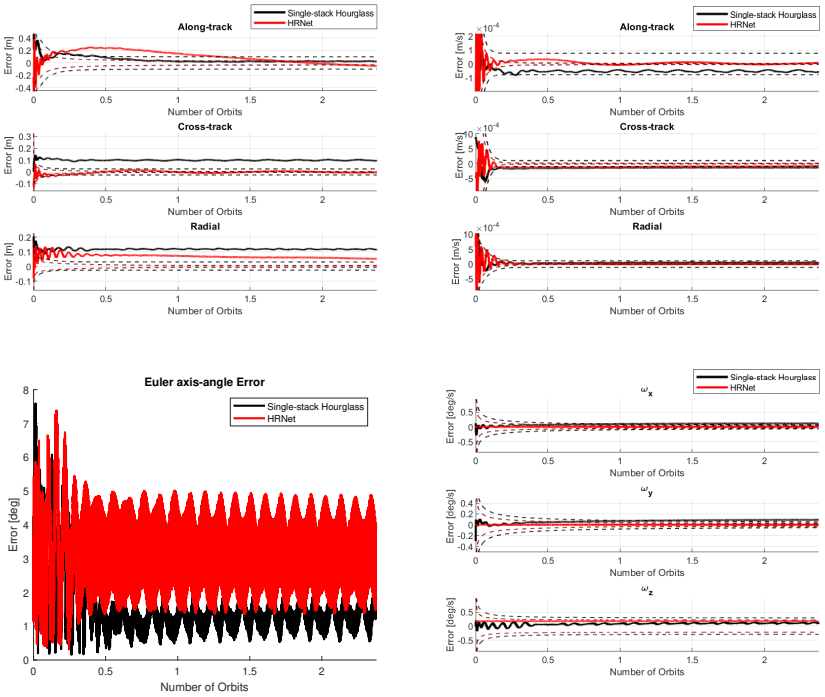



Figure 3.17: Navigation Filter Results - Tightly-coupled MEKF. The dashed lines represent the $1\sigma$ of the estimated quantities.

## 3.8. Chapter Conclusions

This Chapter introduced a novel framework to estimate the pose of an uncooperative target spacecraft with a single monocular camera onboard a servicer spacecraft. A method is proposed in which a keypoint-based CNN is combined with a CEPP$n$P solver and a tightly-coupled MEKF to return an estimate of the pose as well as of the relative translational and rotational velocities. The performance of the proposed method is evaluated at different levels of the pose estimation system, by comparing the detection accuracy of two different CNNs (feature detection step and pose estimation step) while assessing the accuracy and robustness of the selected tightly-coupled filter against a loosely-coupled

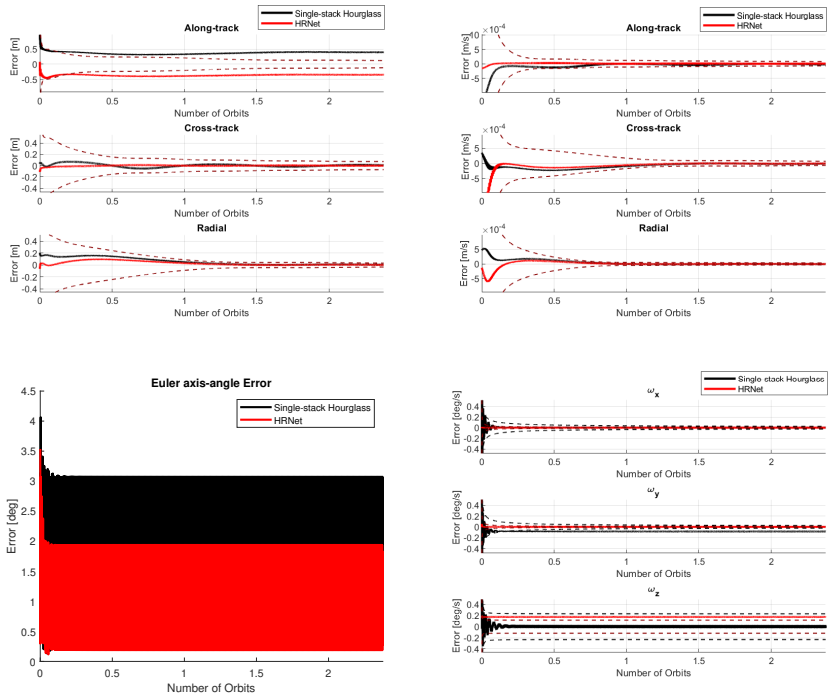



Figure 3.18: Navigation Filter Results - Loosely-coupled MEKF. The dashed lines represent the $1\sigma$ of the estimated quantities.

filter (navigation filter step).

The main novelty of the proposed CNN-based pose estimation system is the introduction of a heatmaps-derived covariance representation of the detected features in a tigthly-coupled, Single-stack Hourglass-based MEKF. At a feature detection level, the performance of the proposed Single-stack Hourglass is compared to the more complex HRNet to asses the feasibility of a reduced-parameters CNN for keypoint detection. Results on the selected test dataset suggest a comparable mean detection accuracy, despite a larger standard deviation of the former network. Notably, this latter aspect is found to decrease the pose estimation accuracy of the Single-stack Hourglass compared to HRNet, despite the adoption of CEPP$n$P to capture features uncertainty. However, important insights are gained at a navigation filter level, delineating two major benefits of the proposed tightly-coupled MEKF. First of all, the capability of deriving a measurement error covariance matrix directly from the CNN heatmaps allows to capture a more representative statistical distribution of the measurements in the filter. Notably, this is expected to be a more complex task if a loosely-coupled filter is used, due to the need to convert the heatmaps distribution into a pose estimation uncertainty through a linear transformation. Secondly, the coupling between the rotational and translational states within the filter guarantees a mutual interaction which is expected to improve the global accuracy of

the filter, especially in the along-track estimate. Besides, the navigation results for the selected vbar scenario demonstrated that the proposed Single-stack Hourglass could represent a valid alternative to the more complex HRNet, provided that its larger detection uncertainty is reflected in the measurement error covariance matrix. Together, these improvements identify an innovative approach to cope with the challenging demand for robust navigation in close-proximity scenarios.

# 4

# Bridging Domain Shift in CNN-based Pose Estimation Systems

## 4.1. Introduction

The navigation results obtained in Chapter 3 highlighted the benefits of adopting a tightly-coupled navigation filter in combination with a keypoint-based CNN for the pose estimation of an uncooperative spacecraft. The simulated V-bar approach scenario around the Envisat spacecraft showed that the capability to derive a measurement error covariance matrix directly from the CNN heatmaps could return a robust navigation estimate and compensate for a lower detection accuracy at image processing level. However, the performance of the proposed pose estimation system was assessed on synthetic images of a simplified rendezvous trajectory. In reality, more challenging actual space imagery could easily jeopardize the navigation performance, especially if a simplified CNN is used. Despite the promising result on the ideal synthetic scenario, the performance of a Single-stack Hourglass on more realistic imagery would in fact be lower than the one of more complex, higher resolution networks such as the HRNet. Remarkably, this is true especially in relation to the domain shift problem introduced in Chapter 1, which states that the performance of a CNN trained from a particular source domain (i.e. synthetic training) can drop unexpectedly when transferred to a different target domain (i.e. actual space imagery). This phenomenon is expected to considerably affect smaller CNNs, due to their reduced number of layers and parameters. In response to this need to guarantee a reliable detection performance of CNN-based systems on realistic imagery, this Chapter investigates the challenges involved in a keypoint-based, HRNet-based pose estimation method when the HRNet is trained on synthetic images and tested on realistic imagery simulated on-ground. The validation of this pose estimation system is ensured by the

Parts of this chapter have been published in Pasqualetto Cassinis et al. (2022a).

introduction of a calibration framework, which returns an accurate reference pose between the target spacecraft and the camera for each lab-generated image, allowing a comparative assessment at a pose estimation level.

Several laboratory testbeds exist to generate images of a target spacecraft's mockup with a monocular camera (Wilde et al., 2019), i.e. the Testbed for Rendezvous and Optical Navigation (TRON) at Stanford University (Kisantal et al., 2020), the GNC Rendezvous, Approach and Landing Simulator (GRALS) at the European Space Research and Technology Centre (ESTEC) (Zwick et al., 2018), the European Proximity Operations Simulator (EPOS) at the German Aerospace Agency (DLR) (Krüger and Theil, 2010), and the Platform-art facility at GMV (Dubanchet et al., 2020). However, only a few detailed calibration procedures were recently described which allow the accurate estimation of the reference pose between camera and target (Park, Bosse, et al., 2021; Valmorbida et al., 2020). Moreover, the calibration of the target spacecraft highly depends on the presence (cooperative target) or not (uncooperative target) of visual markers, as well as on the rendezvous trajectory that shall be recreated (static or moving target). Above all, the challenges in recreating illumination conditions, together with the laboratory constraints on the robot movements, are retained as the main limiting factors in the recreation of realistic rendezvous scenarios. Despite recent efforts aimed at extending the capability to recreate almost any camera-target pose on-ground with highly accurate pose labels (Park, Bosse, et al., 2021), there is still at the time of writing the need to extend the capabilities of on-ground validation setups to allow the recreation of representative rendezvous trajectories.

In relation to the domain shift problem in CNNs, Section 2.3.5 already described how data augmentation and domain adaptation have been used in previous works to leverage the domain shift from synthetic training to real test imagery, highlighting that domain-agnostic data augmentation techniques can generalize the CNN performance from synthetic environments to new domains by using an unrealistic but diverse set of random textures (Geirhos et al., 2019; Jackson et al., 2018; Tobin et al., 2017). Despite promising results on terrestrial applications, the domain shift problem is still a complex and unexplored topic in the space domain, mostly due to the challenges in recreating representative space-like scenarios on-ground. Although recent works investigated the impact of simple training augmentation on the CNN performance using the SPEED lab-generated images (Black et al., 2021a; Park et al., 2019), the laboratory domain was tuned to not differ too much from the synthetic domain. Furthermore, the mockup of the target spacecraft used to generate the lab-images did not differ considerably from the CAD model adopted during synthetic rendering, leading to relatively small domain variations.

Building on the findings described in Chapter 3 and inspired by the promising texture randomization results presented in earlier works (Park et al., 2019), the main objective of this Chapter is to investigate the impact of training data augmentation on the CNN performance on representative space imagery generated on-ground. In order to do so, special focus is put on the recreation of a dedicated calibration pipeline to validate the proposed pose estimation system on representative rendezvous scenarios. The main contributions presented in this Chapter are:

1. To propose a novel CNN training augmentation pipeline focused on texture randomization.
2. To improve the on-ground validation capabilities of the GRALS testbed towards the recreation of representative rendezvous trajectories.
3. To assess the performance of the proposed CNN-based system under challenging domain shifts.

The Chapter is organized as follows. Section 4.2 introduces the proposed on-ground validation framework. The laboratory setup and the calibration procedure are described in Section 4.3-4.4. In Section 4.5, the CNN training, validation and testing phases are detailed, with special focus to the augmentation and randomization pipeline. Next, the pose estimation results are presented in Section 4.6. Finally, Section 4.7 provides the main conclusions and recommendations.



## 4.2. ON-GROUND VALIDATION FRAMEWORK

The on-ground validation pipeline of the proposed pose estimation system is shown in Figure 4.1 and consists of the following main stages:

1. **Calibration procedure and Image Acquisition**: laboratory images of a scaled 1:25 mockup model of the Envisat spacecraft are generated by mounting the camera on a robotic arm which performs a rendezvous trajectory around the mockup. Moreover, the camera is calibrated with respect to the Envisat mockup in order to associate reference labels of the pose between the adopted monocular camera and the mockup for each generated image.
2. **Dataset Generation and CNN Training**: a keypoints-based CNN is trained and validated on augmented datasets. The augmentation is performed by introducing image noise, artificial lights, random background and random textures into synthetically-generated images of the Envisat rendering model.
3. **Online Inference**: the keypoints-based CNN is tested on both synthetic and lab-generated images. The pose is estimated by feeding a P$n$P solver with the detected keypoints as well as with the intrinsic camera parameters and the 3D model of Envisat.
4. **Validation of Pose Estimation Results**: the CNN-based pose estimation results on the lab-generated images are validated against the reference pose labels, derived from the calibrated setup.

### 4.2.1. POSE ESTIMATION SOLVER

Following the results presented in Chapters 2-3, the EP$n$P method followed by Gauss-Newton refinement (Lepetit et al., 2009) is selected to estimate the pose from a set of detected features. This method solves the P$n$P problem in Equation 2.3 in closed-form with the EP$n$P algorithm, and uses the estimated pose as initial guess for an iterative pose refinement.

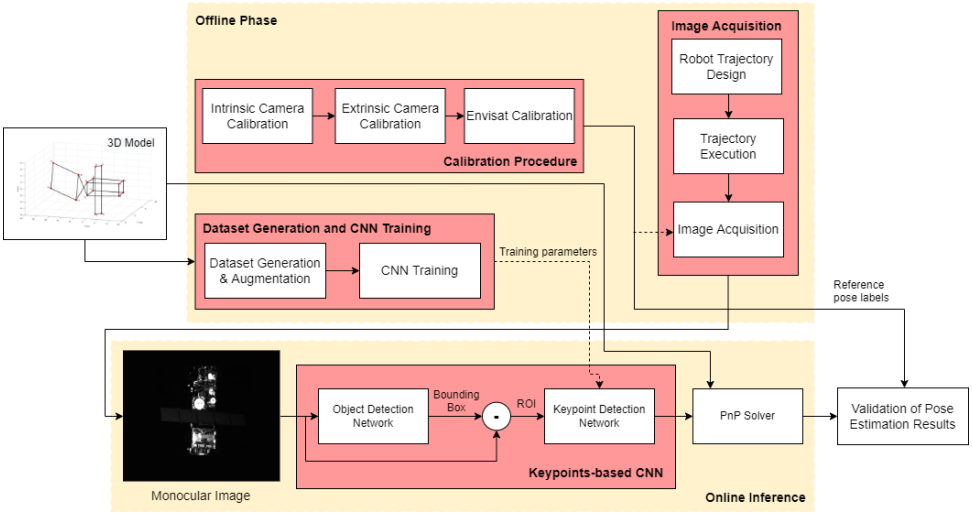



Figure 4.1: Illustration of the proposed on-ground validation of the CNN-based pose estimation system.

## 4.3. THE GRALS TESTBED

The adopted laboratory setup is illustrated in Figure 4.2 and makes use of the GNC Rendezvous, Approach and Landing Simulator (GRALS) testbed of the ORGL facility at ESTEC. The setup consists of the following elements: (a) a 1:25 scaled mockup of the Envisat spacecraft; (b) a Prosilica GC2450 monocular camera; (c) a wall KUKA robotic arm, used to move the Envisat mockup; (d) a ceiling KUKA robotic arm, used to move the camera; (e) the VICON Tracker System (VTS), used to track objects with retro-reflective markers and to provide estimates of their pose with respect to a user-defined reference frame; (f) a lamp mounted on a UR-5 robot, used to recreate the Sun illumination; (g) an external computer providing the software interface between the monocular camera, the VTS and the KUKA robotic arms.

### 4.3.1. VICON TRACKING SYSTEM

The VTS is a highly accurate motion capture system capable of tracking dynamic objects with millimeter accuracy (Merriaux et al., 2017). The system includes a set of 44 calibrated IR cameras, some retro-reflecting spherical markers which can be detected and tracked by the cameras, and a software interface to stream telemetry to the external computer. In the current setup, a subset of 10 cameras is selected such that the total field of view covers the operating volume in which the image acquisition is carried out.

### 4.3.2. KUKA SOFTWARE AND HARDWARE ELEMENTS

The KUKA robotic arms are controlled from the external computer via a Robot Software Interface (RSI) connection. The arms can translate along both ceiling and wall rails and rotate around their six joints, thus guaranteeing a total of 14 degrees of freedom. By default, the command to the robotic arms is represented in terms of end-effector pose

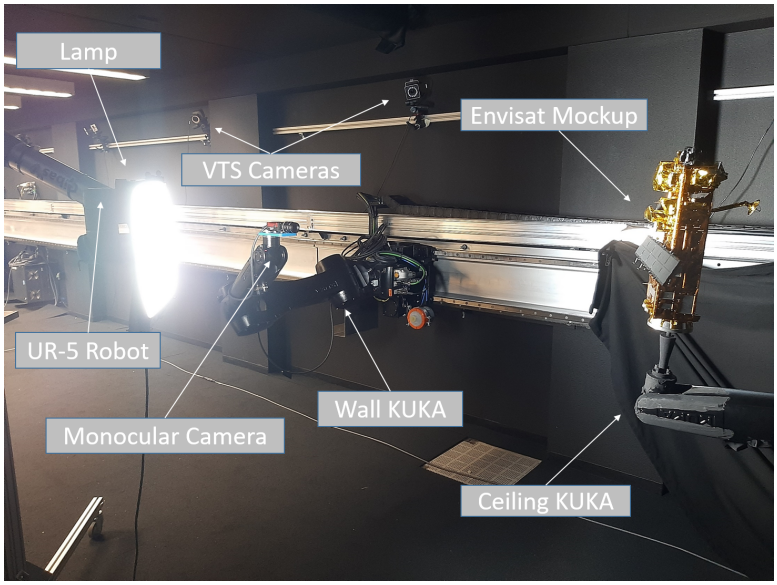



Figure 4.2: GRALS testbed with the scaled 1:25 Envisat mockup mounted on the wall KUKA, the VTS and the monocular camera mounted on the ceiling KUKA. Two of the VTS cameras and the Sun lamp are also shown.



with respect to a pre-defined KUKA base frame. However, the KUKA software allows user-defined *base* and *tool* reference frames, such that any command can be expressed in terms of a selected tool frame pose with respect to a selected base frame.

### 4.3.3. GRALS ILLUMINATION CONDITIONS

In order to recreate a realistic space environment from an illumination standpoint, a movable lamp is mounted on a UR-5 robot and directed towards the target mockup during image acquisition. The lamp is a dimmable, uniform and collimated light source with a spectral response close to 6000 K and an exclusive optical lens which provides high uniformity (±5%), shadow-free backlight illumination[1]. Besides, black curtains are placed around the robots' work zone in order to mask most of the background noise, such as unwanted reflections from the robots' rails.

## 4.4. CALIBRATION FRAMEWORK

The calibration setup consists of the elements described in Section 4.3 and is inspired by the calibration procedure reported in Valmorbida et al. (2020). The objective is to estimate the pose between the monocular camera and the Envisat mockup for each generated image.

### 4.4.1. REFERENCE FRAMES DEFINITION

Referring to Figure 4.3, the following reference frames are defined:

[1] https://www.metaphase-tech.com/backlights/collimated-backlights/

- *LVLH Reference Frame* O: this is the reference frame in which the rendezvous trajectory is expressed (Figure 2.1). Its origin is located at the center of mass of the Envisat mockup and its axes are parallel to the laboratory axes, which represent the radial, along-track and cross-track directions for the simulated relative trajectories.
- *VTS Reference Frame* V: this is the reference frame in which all the objects tracked by VTS are expressed. The frame is defined by a calibration tool consisting of a set of 5 IR markers (Wand calibration object). Notably, the origin and orientation of this frame can be arbitrarily set by the user prior to calibration by placing the Wand object at the desired location.
- *Camera Frame* C: this frame is defined such that the third axis is perpendicular to the image plane and aligned with the optical axis of the camera, with the other two axes planar to the focal plane of the camera.
- *Envisat Body Frame* B: this is a rigid frame oriented with its axis parallel to the sides of the Envisat mockup and centered on the Envisat geometrical center.
- *Ceiling/Wall KUKA end effector frames* CE/WE: these frames are centered at the ceiling and wall KUKA end-effectors, with their third axis perpendicular to the end-effector plate.
- *Markers Object Frame* M: this frame is built from retro-reflective VTS markers attached to a planar surface (not shown in Figure 4.3).

The transformation between each of these frames can be expressed by a roto-translation matrix $\boldsymbol{T}$, which incorporates the relative rotation matrix $\boldsymbol{R}$ and the relative position vector $\boldsymbol{t}$,

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0}_{1\times 3} & 1 \end{bmatrix}. \tag{4.1}$$

### 4.4.2. CALIBRATION PROCEDURE

The purpose of the calibration procedure is to estimate the relative roto-translation matrix $\boldsymbol{T}_B^C$ between the camera frame $C$ and the Envisat body frame $B$ for every monocular image acquired during the desired trajectory. Referring to Figure 4.4, the procedure consists of the following steps:

1. Camera Intrinsic Calibration - Estimation of the Camera Intrinsic Parameters.
2. Determination of the roto-translation matrices $\boldsymbol{T}_V^{WE}$, $\boldsymbol{T}_V^{CE}$ - Calibration of the VTS frame with respect to the Ceiling and Wall KUKA end effector frames and definition of LVLH frame $O$ in both KUKA robots.
3. Estimation of the roto-translation matrix $\boldsymbol{T}_C^V$ - Camera Extrinsic Calibration with respect to the VTS frame.
4. Camera Calibration with respect to the LVLH frame $\boldsymbol{T}_O^C$ - Definition of Camera tool frame $C$ in the wall KUKA.

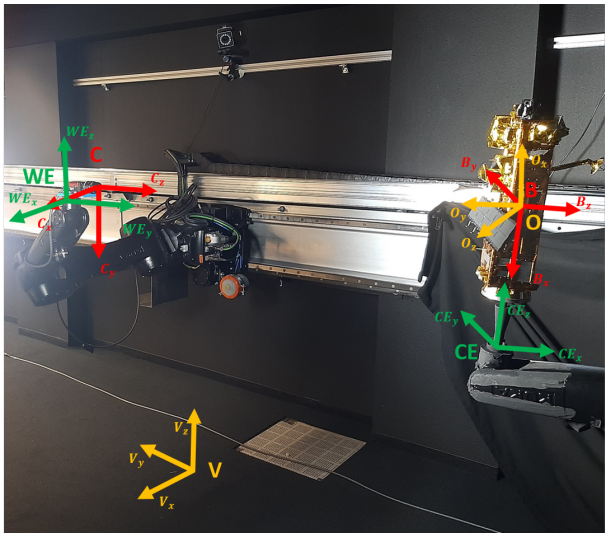



Figure 4.3: Illustration of the reference frames adopted during the calibration procedure. The Wall end-effector (*WE*) and Ceiling end-effector (*CE*) frames are known a-priori, whereas the Camera frame *C* and the Envisat Body frame *B* are unknown and need to be estimated during calibration. The VTS frame *V*, in which the VTS measurements are expressed, can be arbitrarily set by the user. Similarly, the location of the LVLH frame *O* can be chosen based on the constraints on the robot movements for the given laboratory setup. Both the VTS frame *V* and the LVLH frame *O* can be set as tool/base frames in the KUKA environment.

5. Mockup Calibration with respect to the LVLH frame $\boldsymbol{T}_O^B$ - Definition of Mockup Body tool frame *B* in the ceiling KUKA.

6. Computation of the roto-translation matrix $\boldsymbol{T}_B^C$ - Mockup-to-Camera Calibration.

Since the purpose of each calibration step is to define both camera and target tools in their respective end-effector frames, this calibration procedure does not have to be repeated each time a different trajectory is recreated, provided that the same mounting configurations are kept. Also, notice that by calibrating each object with respect to their KUKA end-effectors, it is possible to express them with respect to the common LVLH frame *O*. This is accomplished in order to (i) retrieve representative ground truth relative camera-mockup pose labels for each generated monocular image of the Envisat mockup, and (ii) represent the commanded motion of the robotic arms in terms of camera and Envisat pose with respect to the LVLH frame *O*. This latter aspect is very functional since it is desirable to command the translational and rotational motion in the same reference frame of the intended trajectory.

### 4.4.3. CAMERA INTRINSIC CALIBRATION

The first step of the calibration procedure consists of the estimation of the camera intrinsic parameters needed in Equation 2.3 to solve the P*n*P problem. This is accomplished by taking multiple images of a chessboard with different camera views and using the

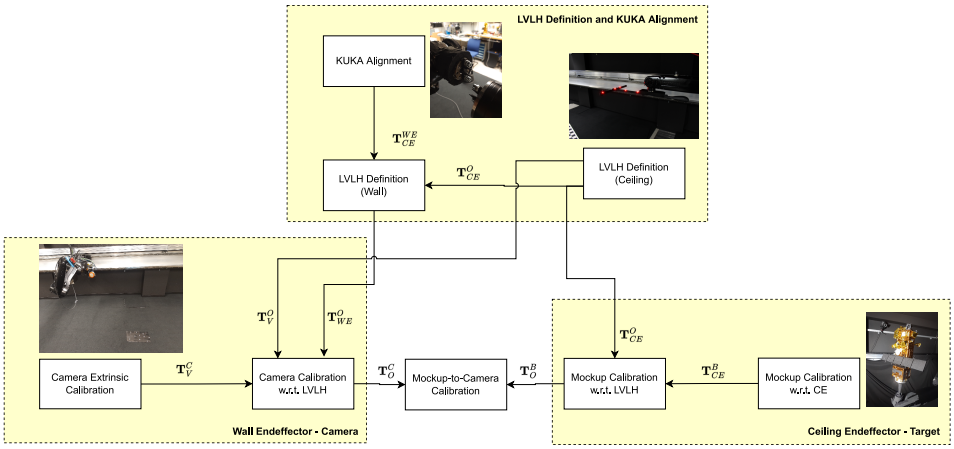



Figure 4.4: Description of the transformations required to compute the final mockup-to-camera pose. The LVLH definition in both the wall and ceiling KUKA is done in order to define the LVLH frame as base frame in both robots.

*estimateCameraParameters* [2] Matlab built-in function. This function estimates the intrinsic parameters $[f_x, f_y, C_x, C_y]$ and the distortion coefficients of a monocular camera, whilst also returning the images used to estimate the camera parameters and the standard estimation errors for the single camera calibration.

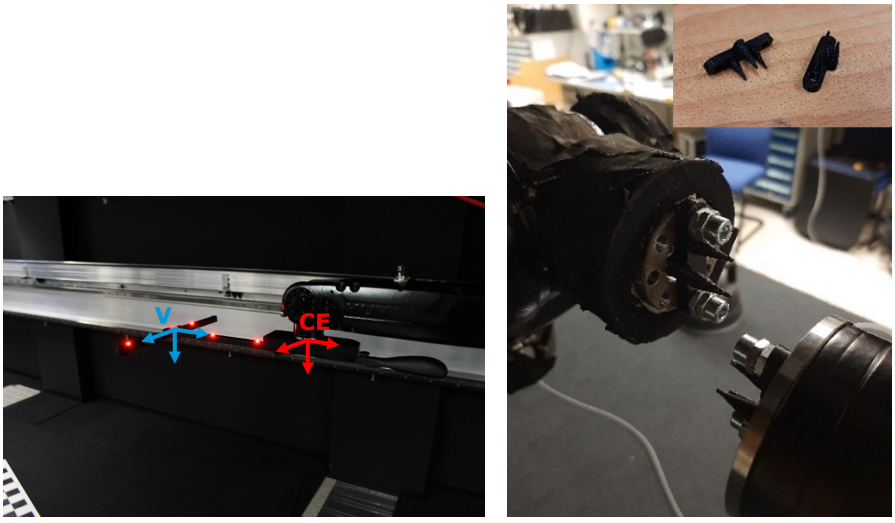
### 4.4.4. VTS FRAME CALIBRATION AND DEFINITION OF LVLH FRAME *O*

The calibration of the VTS is an essential step towards the overall calibration of the GRALS setup. This calibration is performed (i) to define the LVLH frame *O* in which the rendezvous trajectory will be represented, and (ii) to express the camera frame *C* in the Wall end-effector frame *WE* after the camera extrinsic calibration. To do so, the first step consists in defining the VTS frame *V* in the Ceiling end-effector frame *CE* (Figure 4.5a). This is done by mounting the VTS's Wand calibration object onto the CE and exploiting the knowledge of the geometry of the mount. At this stage, the LVLH frame *O* is constructed by matching it to the V frame ($O = V$), in order to ease frames transformations. Once the VTS frame is expressed in the CE frame, the Wand object is removed from the CE, and alignment pins are mounted on both the CE and the WE to align the two KUKAs with sub-millimeter accuracy (Figure 4.5b). As a result, the *V* frame can be expressed in both end-effector frames interchangeably.

### 4.4.5. CAMERA EXTRINSIC CALIBRATION

A high-level schematic of the extrinsic camera calibration procedure is illustrated in Figure 4.6. The first task is to recreate a planar object M by placing some retro-reflective markers onto a planar surface. Based on similar setups (Valmorbida et al., 2020), 10 markers were used to recreate the object M.
Next, the planar object M is moved in order to generate pictures of the retro-reflective

[2] www.mathworks.com/help/vision/ref/estimatecameraparameters.html




(a) VTS's Wand mounting on ceiling's end effector

(b) Alignment of the two KUKA end-effectors with the use of alignment pins

Figure 4.5: VTS frame (V) definition with respect to the Ceiling end-effector frame (CE) and alignment of the two KUKA end-effectos with the use of alignment pins.

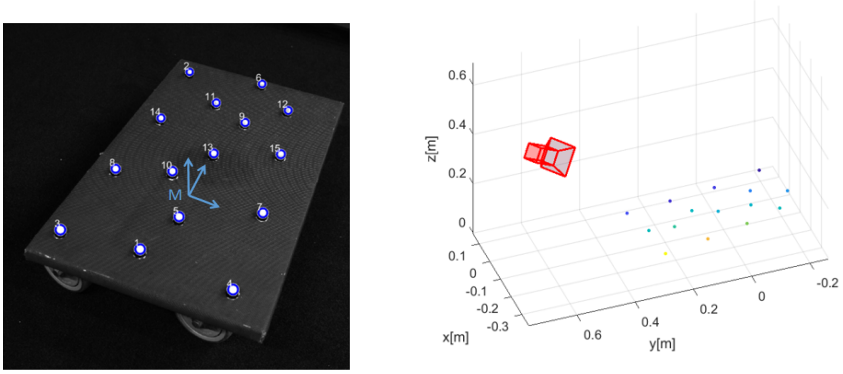
markers under different camera views. The pixel location of each marker is then extracted by using the Matlab built-in Circular Hough Transform (CHT) algorithm. A manual 2D-3D point correspondence is performed in order to associate each detected marker with its three-dimensional location in the $M$ frame. At this stage, the EP$n$P algorithm is used to solve the P$n$P problem and obtain an estimate of the roto-translation between the camera frame $C$ and the VTS frame $V$, exploiting the knowledge of the pose of the markers object M with respect to the VTS frame.

Subsequently, multiple images of the object $M$ are taken with different camera views, and the CHT is applied to each of them to extract the pixel location of the retro-reflective markers. A total of 15 images is chosen in order to capture enough variation in the camera view of the markers object. For each frame, the 2D-3D point correspondence can be made by using the initial estimate of $\boldsymbol{T}_C^V$. The P$n$P problem can then be solved by means of a non-linear least squares solver, by minimizing the following sum of squares (Valmorbida et al., 2020):

$$\sigma_1(\boldsymbol{x}) = \sum_{k=1}^{N_p} \sum_{j=1}^{N_m} \left\| \boldsymbol{p}_{f,i}(k) - \boldsymbol{\pi}\left(\boldsymbol{m}_{f,i}^V(k), \boldsymbol{T}_C^V\right) \right\| \tag{4.2}$$

$$\boldsymbol{\pi}\left(\boldsymbol{m}_{f,i}^V(k), \boldsymbol{T}_C^V\right) = \left( \frac{x_{f,i}^C}{z_{f,i}^C} f_x + c_x, \frac{y_{f,i}^C}{z_{f,i}^C} f_y + c_y \right) \tag{4.3}$$

$$\boldsymbol{m}_{f,i}^C = \begin{bmatrix} x_{f,i}^C & y_{f,i}^C & z_{f,i}^C \end{bmatrix}^T = \boldsymbol{R}_V^C \boldsymbol{M}_{f,i}^V + \boldsymbol{t}_C^V \tag{4.4}$$




(a) (b)

Figure 4.6: Estimation of the roto-translation between the camera frame C and the markers frame M. The markers detection by the CHT algorithm (a) is shown beside the estimated roto-translation of the camera with respect to the the M object (b).

where $N_m$ is the number of fiducial markers, $N_p$ is the number of frames, $\boldsymbol{m}_{f,i}^{M}$ represents the location of the i$^{th}$ marker in the VTS frame $V$, and $\mathbf{r}_i^C = \begin{bmatrix} x_i^C & y_i^C & z_i^C \end{bmatrix}^T$ is the position of each feature with respect to a camera with focal length $\boldsymbol{f} = \begin{bmatrix} f_x & f_y \end{bmatrix}$ and principal point $\boldsymbol{c} = \begin{bmatrix} c_x & c_y \end{bmatrix}$ (Eqs. 3.9-3.10). The output of the minimization is a refined estimate of $\boldsymbol{T}_C^V$, which is used to reproject the 3D retroreflective markers on the image plane and compute the deviation from the correct 2D location of each marker. Figure 4.7 shows the reprojection error across the whole set of images of the markers object M. The pixel error can be represented by a distribution with $\boldsymbol{\mu} = [0.14, -0.15]\,px$ and $\boldsymbol{\sigma} = [1.6, 2]\,px$. Overall, the pixel error deviation does not exceed 0.08% of the image size. This is in agreement with the extrinsic calibration results obtained by Valmorbida et al. (2020).

### 4.4.6. MOCKUP CALIBRATION

The calibration of the Envisat mockup consists in estimating the pose of the Envisat body frame $B$ with respect to the Ceiling end effector frame $CE$ (Figure 4.8). Thanks to the adopted design for the mount, which guarantees a unique fixation of the mockup onto the CE, this transformation can be derived directly from the CAD geometry of the mount and the location of the $B$ frame with respect to the mockup mounting interface. Although in principle a dedicated mockup calibration via retro-reflective VICON markers should return more accurate results, the challenges in reconstructing the transformation from the markers frame to the body frame $B$ is currently considered a limiting factor. Specifically, the large number of instruments located on the target and the uneven surface of the Multi-Layer insulation (MLI) prevent from accurately mounting the markers.

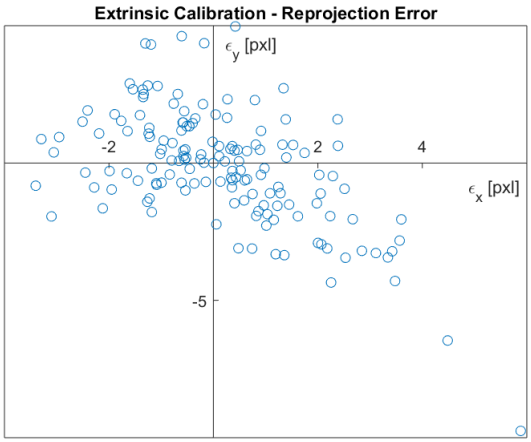



Figure 4.7: Reprojection error after camera extrinsic calibration.

### 4.4.7. GLOBAL CALIBRATION ERROR ANALYSIS

Overall, the calibration steps described in the previous sections are essential to estimate the desired Mockup-to-Camera transformation $\boldsymbol{T}_B^C$. Each of these steps is characterized by individual calibration inaccuracies that contribute to the global error of the calibration setup $E_{\text{Cal}}$:

$$E_{\text{Cal}} = E_{\text{Cal}}(E_{\text{Int}}, E_{\text{Ext}}, E_{\text{VTS}}, E_{\text{Al}}, E_{\text{KUKA}}, E_{\text{Env}}), \tag{4.5}$$

in which:

- $E_{\text{Int}}$, $E_{\text{Ext}}$ represent the reprojection error due to the intrinsic and extrinsic camera calibration with respect to the VTS frame V;
- $E_{\text{VTS}}$ represents the VTS pose error due to inaccuracies in the detection of the retro-reflective markers by the VTS cameras;
- $E_{\text{Al}}$, $E_{\text{KUKA}}$ represent the pose error due to the robots alignment step and due to the intrinsic KUKA inaccuracies, respectively;
- $E_{\text{Env}}$ represents the reprojection error due to inaccuracies in the Envisat-to-CE mount.

Notably, deriving a quantitative global calibration accuracy for each calibration term is complicated by the fact that some of these accuracies are expressed in terms of reprojection error onto the image plane, whereas others are expressed in terms of pose error. To cope with this limitation, the impact of the global calibration error on the estimation error of the transformation $\boldsymbol{T}_B^C$ is assessed by monitoring the reprojection error $\epsilon$ of the mockup's corners on a small subset of five of the generated monocular images of the

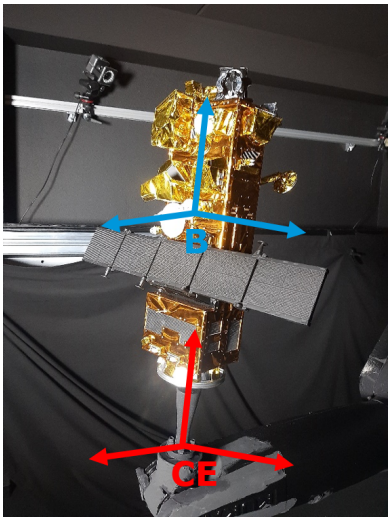



Figure 4.8: Illustration of the mounting of the Envisat mockup on the CE.

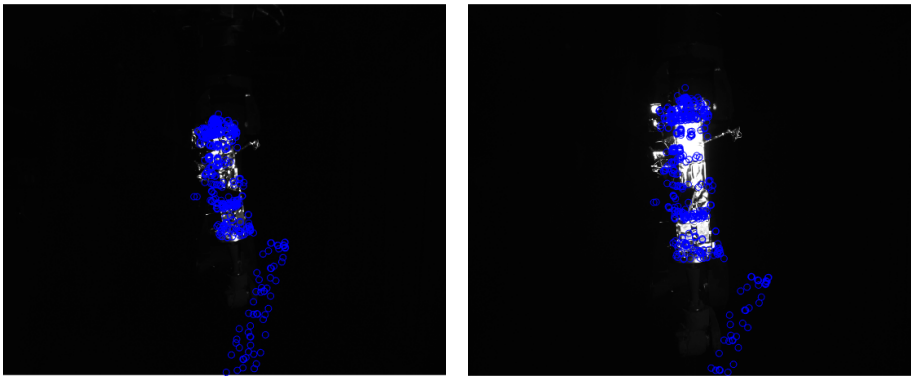
target spacecraft. This reprojection can be obtained by manually selecting the visible corners in each image of the subset and by comparing their pixel location with the reprojection based on each 3D point from the estimated camera intrinsic parameters and the calibrated pose $\boldsymbol{T}_B^C$ (Equation 2.3). Table 4.1 lists the error contribution of each calibration step together with the final mean reprojection error on the selected subset. Notice that the Envisat mounting error $E_{\text{Env}}$ could not be quantified due to the unknown relation between the mounting inaccuracies and the resulting mockup pose inaccuracy. Besides, the entire point cloud, obtained from the CAD model of the mockup, can be reprojected onto each generated image to get a qualitative measure of the calibration accuracy. Figure 4.9 illustrates two representative examples of the point cloud reprojection in different relative ranges, together with the mean reprojection error derived from the visible corners. Overall, the same order of magnitude of the reprojection error is observed for the remaining images of the subset, sampled across the trajectory to cover different camera-target ranges.

The proposed calibration procedure exhibits a larger total reprojection error, when compared to the sub-pixel results obtained by Park, Bosse, et al. (2021) with a more dedicated hand-eye calibration (Tabb and Yousef, 2017) of Stanford's TRON facility. However, the calibration error of TRON was minimized and computed on a subset of very close-range poses of the target spacecraft. As such, an increase in the reprojection error for larger relative ranges is expected. Yet, the uncertainty in the Envisat mounting suggests that $E_{\text{Env}}$ is the main contributor to the larger calibration error observed in the proposed GRALS setup. Nevertheless, it is expected that the pose errors resulting from the reprojection offsets can still guarantee a representative ground truth for the intended on-ground validation.

Table 4.1: Global calibration error analysis. Note that the Envisat mounting error $E_{\text{Env}}$ could not be quantified due to the unknown relation between the mounting inaccuracies and the resulting mockup pose inaccuracy. As described in the text, the total reprojection error is computed by comparing the reprojected corners of Envisat with the visible corners in a subset of 5 representative images. As such, the total reprojection error is not computed as a function of each individual error.

| **Calibration Error** | **Value** | **Description** |
|---|---|---|
| $E_{\text{Int}}$ | 0.22 px | Reprojection error |
| $E_{\text{Ext}}$ | (0.14,-0.15) px ± (1.6,2) px | Reprojection error |
| $E_{\text{VTS}}$ | <1 mm | Markers detection error |
| $E_{\text{Al}}$ | 0.1 mm/0.02° | KUKA accuracy |
| $E_{\text{KUKA}}$ | 0.1 mm/0.02° | KUKA accuracy |
| $E_{\text{Env}}$ | N/A | Mounting error |
| $E_{\textbf{Cal}}$ | **18.7 px** | **Total Reprojection error** |





(a) $\boldsymbol{t}_B^C$ = 2.8 m - $\epsilon$ = 19.5 px (b) $\boldsymbol{t}_B^C$ = 2 m - $\epsilon$ = 18 px

Figure 4.9: Reprojection of the Envisat point cloud onto the image plane for two representative Mockup-to-Camera poses. The mean reprojection error $\epsilon$ is computed by manually selecting the visible corners of the mockup and by comparing their pixel location with the reprojected values derived from calibration.

### 4.4.8. RENDEZVOUS TRAJECTORY GENERATION

Once the GRALS testbed is fully calibrated, any relative trajectory between the monocular camera and the target mockup can be recreated by commanding the two KUKA robotic arms in terms of camera/mockup tool frames with respect to the LVLH frame $O$. To comply with the physical constraint of the robotic arms, a rectilinear approach in-line with the flight path of the servicer spacecraft towards the target spacecraft (so called *V-bar* approach) is considered, as this typically occurs during the final stages of close-proximity operations in rendezvous and docking missions (Tatsch et al., 2006; Wieser et al., 2015). This assumption is justified by the fact that the CNN performance on lab-generated images shall be first validated on simplified relative trajectories, before assessing its performance under more complex relative geometries. Following the same line of reasoning, the relative attitude of the target is simplified by recreating a spinning rotation of around 3.5 deg/s along the main longitudinal body axis, superimposed with precession. The magnitude of the Envisat rotation complies with past optical observation data (Hou-Yuan and Chang-Yin, 2018), whereas the direction of rotation is chosen based on the constraints in the robotic arm movements. Moreover, relative distances of 4 m down to 1 m are recreated in the lab which correspond to relative distances in the range of 100 m - 25 m for the full-scale target spacecraft. Lastly, the UR-5 robot is used next to the two KUKA robotic arms to control the lamp at 40°, 60°, and 90° Azimuth angles with respect to the Envisat mockup. The location of the lamp is kept fixed throughout the trajectory, but it is changed at the end of each execution in order to execute the same V-bar approach under different illumination conditions. Figure 4.10 shows a sample pose under varying Azimuth illumination angles. To guarantee consistency throughout all these illuminations, a LUX-meter is used to ensure that the same light irradiance of 1366 W/m$^2$ (typical of LEO orbits) is kept while changing the Azimuth angle. Although the exact distribution of the irradiation across different wavelength is not monitored, this ensures that the light irradiance on the target surface does not change for different illumination angles. Notably, the use of a lamp as opposed to a more diffusive illumination guarantees worst-case reflections on the target satellite. As a result, the CNN performance can be stress-tested on worst-case illumination scenarios which differ from the synthetic renderings.

It is important to mention that, although a realistic close-proximity approach would undergo varying illumination angles over time due to the motion of the Sun with respect to the LVLH frame, the current assumption of a fixed light source during the approach is justified by the relatively short duration of the relative trajectory. At the same time, the selected Azimuth range is considered representative of the currently planned ADR missions, since close-proximity approaches at larger illumination angles are typically avoided with careful mission design.

## 4.5. CONVOLUTIONAL NEURAL NETWORK

As described in Section 2.3.5, recent advances in keypoint-based CNNs demonstrated that by using parallel sub-networks across multiple resolutions, rather than multi-resolution serial stages, a CNN is able to maintain a richer feature representation, facilitating more accurate and precise heatmaps. For this reason, the HRNet architecture (Section 3.3.1)

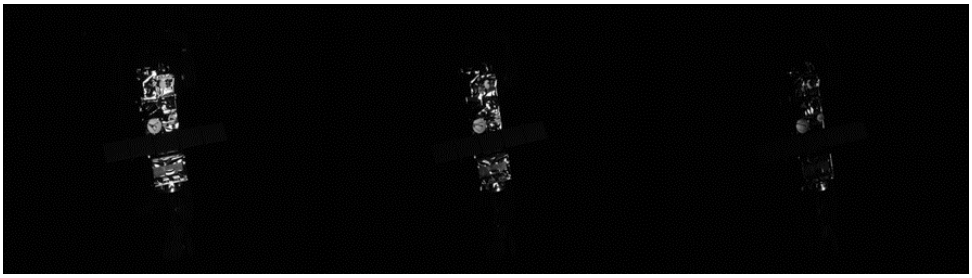


Figure 4.10: Example of different illumination conditions for a sample pose. Azimuth illumination angles of 40° (left), 60° (center), and 90° (right) are shown.



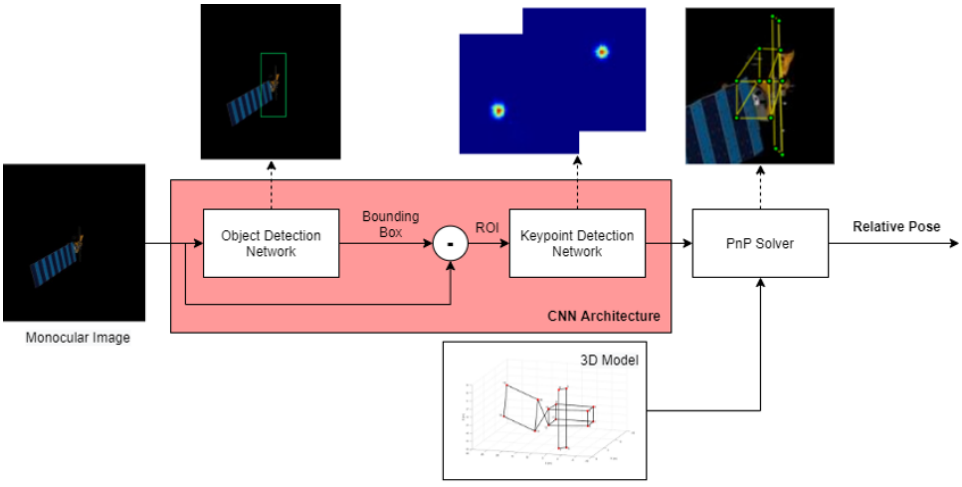



Figure 4.11: Proposed CNN architecture and interface with the P*n*P solver.

currently represents the state-of-the-art in keypoint detection, and is chosen in the proposed pose estimation system following its recent performance results on the SPEED dataset (Chen et al., 2019). Figure 4.11 illustrates the adopted CNN architecture together with its interface with the P*n*P solver.

### 4.5.1. AUGMENTATION PIPELINE

In Figure 4.12, the first step of the proposed pipeline for the datasets augmentation and randomization consists in generating ideal synthetic images of the Envisat 3D model in Cinema 4D©. The same synthetic dataset described in Section 3.3.3 is adopted with the same camera parameters (Table 4.2). Next, a randomization pipeline is introduced which adds the following effects to the rendering:

- Texture randomization. This is performed in order to increase the CNN robustness against texture variations between the synthetic and lab models of Envisat. The randomization is achieved in two different ways, by either adding a shader to each

Table 4.2: Parameters of the camera used to generate the synthetic images in Cinema 4D©.

| Parameter | Value | Unit |
|---|---|---|
| Image resolution | 256×256 | pixels |
| Focal length | $3.9{\cdot}10^{-3}$ | m |
| Pixel size | $1.1{\cdot}10^{-5}$ | m |



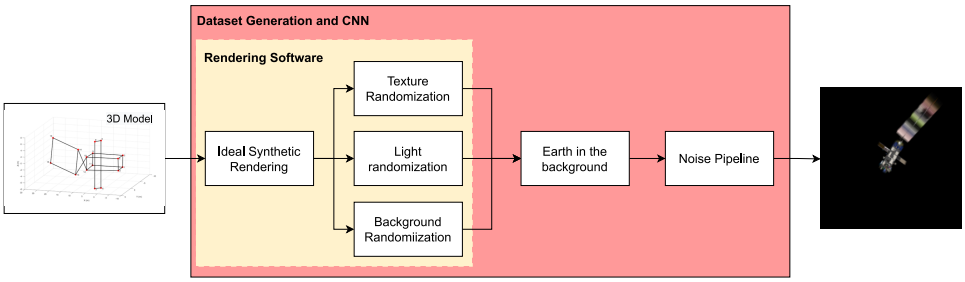



Figure 4.12: Dataset Augmentation Pipeline.

material in order to noise the textures, or by directly shuffling the textures of the materials. Besides, the reflectance of each material of the rendering model is also randomized, in order to increase the variability of the material properties across the target body.

- Light randomization. Four additional lights are introduced in random locations, aside from the main Sun illumination, in order to increase the CNN robustness against the illumination conditions recreated in the laboratory setup.
- Background randomization. Random scenes are used as image background in order to increase the CNN robustness against the laboratory environment. Specifically, external disturbance sources in the lab are likely to return non-zero pixel values in the image background, leading to inaccurate CNN detections if the training dataset would lack of background augmentation.

Remarkably, the proposed texture randomization differs from most of the implementations described in Section 4.1 in that it takes place *before* the actual rendering, and not in post-processing. As a result, the randomization can be performed directly on the actual spacecraft materials and textures without jeopardizing the target body shape. This latter aspect could happen when random texture patterns are superimposed to the target image after rendering. Furthermore, the inclusion of both texture and light augmentation aims at generalizing the training domain to the GRALS testbed's illuminations whilst improving the CNN robustness against previously unseen textures of the target mockup.

Following the Cinema4D© rendering already described in Section 3.3.3, an additional pipeline is used to further augment the generated images and improve the network robustness towards realistic space images of the target. This is performed by introducing the Earth in the background in some of the images and by corrupting the images with the following noise models:

- Gaussian, shot, impulse and speckle noise
- defocus, motion and zoom blurs
- spatter, color jitter and random erase.

Table 4.3 lists all the augmentation techniques together with the number of generated images, whereas Figure 4.13 shows four representative examples of the adopted data augmentation techniques. A total of 24,400 images were rendered and further split into training (70%), validation (15%) and test (15%) datasets. These percentages were selected based on other augmentation pipelines (Black et al., 2021c).



Table 4.3: Augmentation Breakdown. The randomization in the last augmentation step refers to both random lights, textures, and background.

| Description | number of Images |
|---|---|
| No augmentations | 1000 |
| Random lights | 550 |
| Random lights & textures | 2000 |
| Random lights & background | 350 |
| All randomizations & Noise & Earth | 20,500 |
| **Total** | **24,400** |

### 4.5.2. TRAINING, VALIDATION AND TEST

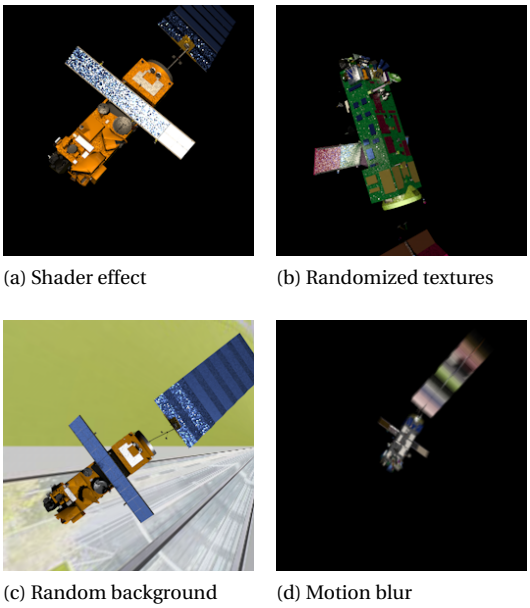
The same network configurations described in Section 3.3.3 are used during training and validation, with the Adam optimizer used with a cosine decaying learning rate with initial value of $10^{-3}$ and decaying factor of 0.1. Then, the network performance is assessed with the augmented synthetic test dataset as well as with the GRALS-generated images. A Tesla P100-PCIE-16GB GPU is used for both training and testing.

## 4.6. RESULTS

In this section, the pose estimation results are presented for the V-bar trajectory images generated at ESTEC's GRALS testbed. The error metric introduced in Section 3.7 is used, namely:

$$E_T = \left\| \boldsymbol{t}^C - \hat{\boldsymbol{t}}^C \right\| \tag{4.6}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_s & \boldsymbol{\beta}_v \end{bmatrix} = \boldsymbol{q} \otimes \hat{\boldsymbol{q}} \tag{4.7}$$



(a) Shader effect (b) Randomized textures

(c) Random background (d) Motion blur

Figure 4.13: Output examples of the randomization pipeline.

$$E_R = 2\arccos(|\beta_s|) \tag{4.8}$$

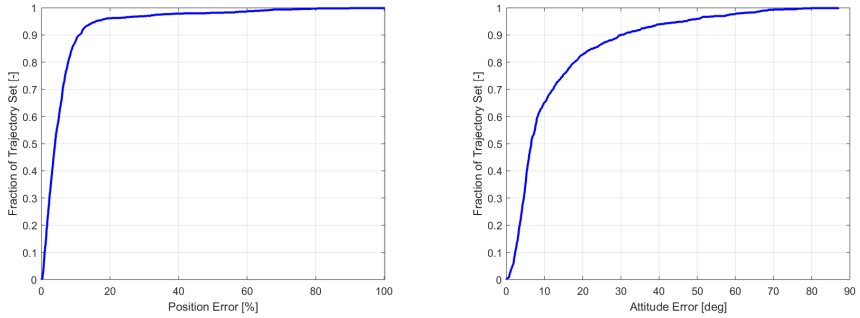
where the translational error is expressed as the norm of the difference between the estimated relative position $\hat{\boldsymbol{t}}^C$ and the ground truth $\boldsymbol{t}_C$, and the rotational error is expressed in terms of the Euler axis-angle error between the estimated quaternion $\hat{\boldsymbol{q}}$ and the ground truth $\boldsymbol{q}$. To categorize the pose estimation error, the following definitions are introduced:

- High accuracy: $E_T < 5\%$, $E_R < 2°$,
- High/medium (Medium) accuracy: $E_T < 10\%$, $E_R < 5°$,
- High/medium/low (Low) accuracy: $E_T < 10\%$, $E_R < 10°$,

in which the position error is expressed as a percentage of the ground truth relative position $\boldsymbol{t}^C$. Moreover, if the number of keypoints within the defined detection threshold falls below the minimum number of features required by the EP$n$P to solve for the pose, no pose is returned.

### 4.6.1. HIGH EXPOSURE, 40° ILLUMINATION AZIMUTH

Table 4.4 lists the categorized pose estimation results as a percentage of the High Exposure, 40° Illumination Azimuth V-bar trajectory images. As can be seen, 59% of the trajectory images are characterized by position errors $E_T < 10\%$ and attitude errors $E_R < 10°$. Moreover, medium and high accuracies are achieved in 31% and 3% of the images, respectively.




(a) (b)



Figure 4.14: Cumulative distribution function for the position (a) and attitude (b) errors. As can be seen, an initial steep increase in both curves highlights that most of the images are characterized by relatively low pose errors.

Furthermore, Figure 4.14 shows the cumulative distribution function for both the attitude and position errors across the V-bar trajectory. This function is convenient in that it captures which fraction of the trajectory images returns a certain pose estimation accuracy. Overall, these results highlight a remarkable performance of the proposed pose estimation system. Despite the limitations in the achievable calibration accuracies reported in Section 4.4.7, the results demonstrate that a CNN trained on augmented, purely synthetic images can adapt to a previously unseen domain, and perform accurate keypoints detections. Specifically, the inclusion of a texture randomization step within the data augmentation pipeline indicates that the CNN focuses on the shape of the target rather than on its textures. This improves the detection robustness against illumination conditions and material reflections that were not part of the training dataset.

To help analyze the CNN detection performance prior to pose estimation, Figure 4.15 illustrates four representative CNN detections for each pose accuracy category. First of all, a scenario characterized by an adverse MLI reflection is shown for which no pose estimate is returned. These unfavorable reflections are very challenging to handle by the CNN, resulting in a highly scattered and inaccurate keypoints detection. Next, the pose estimate scenarios are displayed. Notably, a lower detection accuracy can be inferred for the upper corners in the high and medium accuracy estimates. This is deemed to be a direct consequence of instruments occlusion, which is not properly captured in the training dataset due to the differences between the target mockup and the rendering model. Furthermore, a low accuracy in the detected SAR antenna keypoints can be observed in the low accuracy scenario. The SAR antenna corners are generally easier to detect than other target keypoints, mostly due to a higher similarity with the rendering model and a lack of adverse MLI reflection. As a result, they are retained as the main contributors to high and medium pose accuracies, leading to lower pose accuracies when not accurately detected.

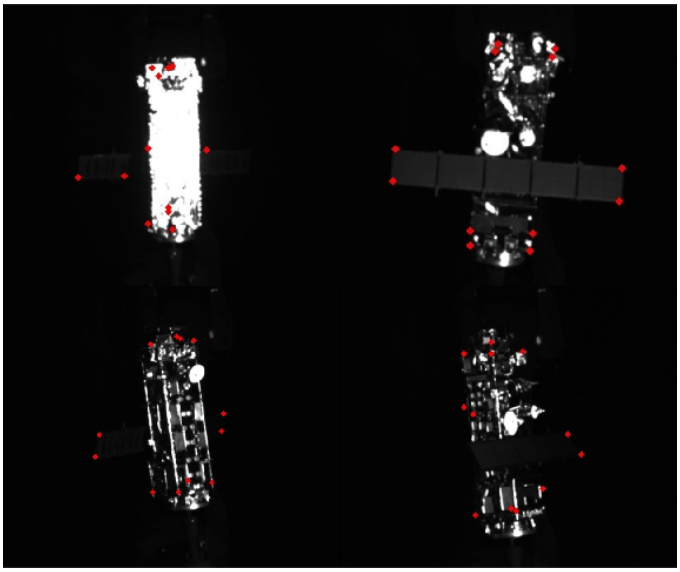


Figure 4.15: Upper left: No pose due to adverse MLI reflection. Upper right: High accuracy. Lower left: Medium accuracy due to instruments occlusion on upper corners. Lower right: Low accuracy due to improper SAR corners detection and offset in the lower corners.

To further investigate the overall performance of the proposed system, the pose estimation results are extended to the entire V-bar trajectory. Firstly, Figure 4.16 shows the camera boresight component of the estimated position against the ground truth relative distance. Overall, it can be seen that the estimation follows the true value with error peaks scattered throughout the trajectory. Notably, a larger number of outliers is observed for mid-far ranges, suggesting an increase of pose estimation error with distance. Next, Figure 4.17 illustrates the pose estimation results after averaging with a moving mean with a window size $k = 100$. This is done in order to capture the relation between the mean estimation error and the relative distance between the monocular camera and the target. As a validation, both the position and attitude errors exhibit the typical trend observed in monocular pose estimation systems (Kisantal et al., 2020; Park et al., 2019; Sharma and D'Amico, 2015), with larger mean position errors at larger distances and comparable attitude errors unless the target is less than 45 m away from the camera. Furthermore, the obtained mean attitude errors are comparable to the ones obtained by other pose estimation systems on the lab-generated images of the SPEED dataset (Kisantal et al., 2020). This is remarkable since, although the SPEED dataset includes 300 images under several poses, the adopted illumination source consisted of light boxes resembling the Earth albedo in no-eclipse scenarios. This is an illumination condition far less extreme than a direct, high intensity Sun lamp, due to the patterned flare introduced by the sun lamp and intense surface glow due to high reflectivity and overexposure of the camera. As a result, a smaller domain adaptation is required from the synthetic SPEED dataset compared to the GRALS trajectory images.

Table 4.4: Pose Estimation Results for the high exposure, 40° Azimuth V-bar trajectory. Position results are scaled from the ORGL setup to the real orbital distances by accounting for the real dimensions of the target spacecraft. The remaining 6% of the trajectory images are characterized by pose errors above the threshold set for the low accuracy.

| Scenario | No Pose | High accuracy | Medium accuracy | Low accuracy |
|---|---|---|---|---|
| High exp. 40° Az. | 1% | 3%<br>$\vec{E}_T = 0.2$ m<br>$\vec{E}_R = 0.8°$ | 31%<br>$\vec{E}_T = 0.8$ m<br>$\vec{E}_R = 3°$ | 59%<br>$\vec{E}_T = 1.4$ m<br>$\vec{E}_R = 4.6°$ |



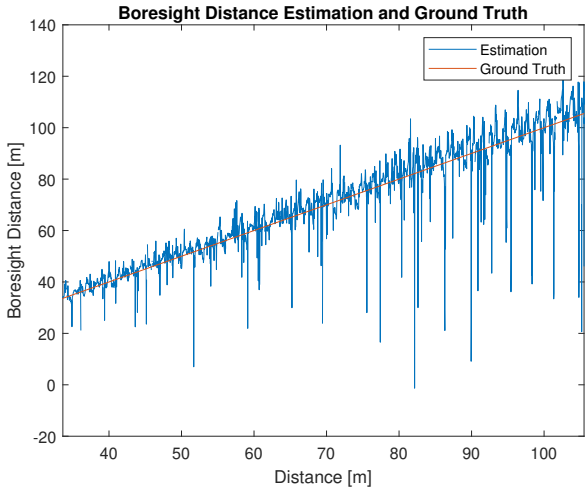
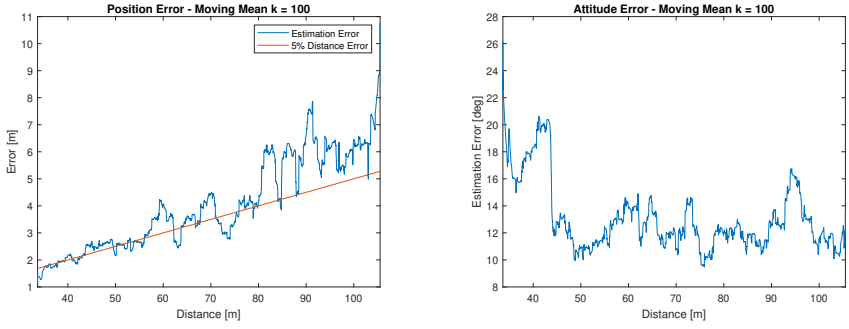



Figure 4.16: Boresight estimation from the CNN+EP*n*P pipeline compared to the ground truth pose from calibration.




(a) (b)

Figure 4.17: Moving average trends of the estimated relative position (a) and attitude (b).

Table 4.5: Pose Estimation Results for the low exposure, 60°-90° Azimuth V-bar trajectory.

| Scenario | No Pose | High accuracy | Medium accuracy | Low accuracy |
|---|---|---|---|---|
| Low exp. 60° Az. | 38% | 0.3% | 3% | 5% |
| Low exp. 90° Az. | 75% | 0 | 0.5% | 2% |

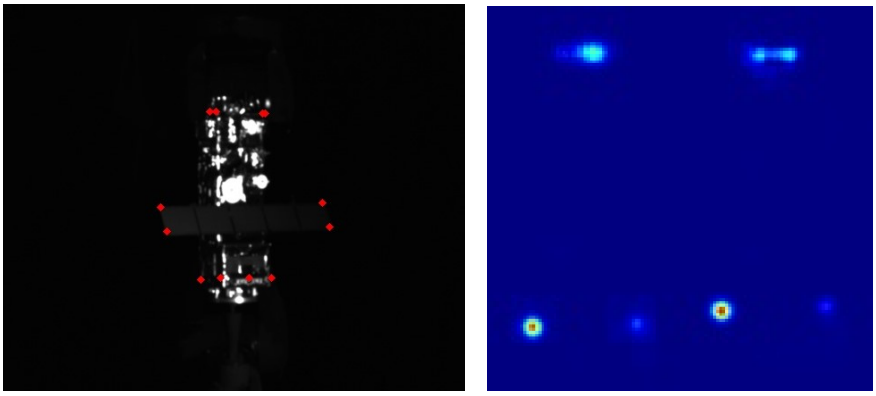


### 4.6.2. LOW EXPOSURE, 60°-90° ILLUMINATION AZIMUTH

Table 4.5 lists the categorized pose estimation results as a percentage of the low Exposure, 60°-90° Illumination Azimuth V-bar trajectory images. As anticipated in Section 4.4.8, these trajectories are characterized by more adverse illumination conditions as well as by a much lower exposure of the monocular camera, in order to stress-test the CNN performance on extreme scenarios. As can be seen, the pose estimation accuracy drops considerably compared to the results observed in Table 4.4, indicating that the adopted training data augmentation is not enough to bridge this synthetic-lab domain gap. However, the severe illumination conditions in these two scenarios suggest that the main cause of a larger domain adaptation could be traced back to the randomization of the main light source locations recreated during training. In other words, the extremely low pose accuracies are not expected to be a direct result of an insufficient texture randomization, and further improvements in the CNN training shall aim at extending the illumination scenarios.

### 4.6.3. POSE ERROR ANALYSIS

The pose estimation analysis in Section 4.6.1 provided important insights on the performance of the CNN in the high exposure V-bar scenario, proving its capability to return satisfactory pose estimates for over half of the images. Yet, it is also important to investigate the scenarios in which the pose estimate considerably drifts from the ground truth, i.e. the large errors observed in Figure 4.16. Although the majority of the pose outliers are related to a poor keypoints detection, there might be cases in which the large estimation error stems from solving the P$n$P problem rather than from the CNN detection step. Figure 4.18a illustrates a representative example: as can be seen, the CNN performs an accurate detection of most of the keypoints. However, the attitude error associated to this detection amounts to $E_T > 50°$. Notably, the fact that the position error is relatively small suggests that the estimation algorithm is correctly locating the target but confusing its orientation. Generally, this is an indication that there could be a wrong 2D-3D correspondence before solving the P$n$P problem (Equation 2.3). Specifically, a wrong correspondence would happen if the CNN suddenly switches the keypoints detection order, as this would associate the wrong 2D features to the 3D model points. Since heatmaps are a good indicator of the CNN confidence on each detection, it could be possible to correlate the heatmaps shape and intensity with a potential features switch. Following this line of reasoning, Figure 4.18 shows a representative example in which



(a) CNN keypoints - frame k+1 (b) CNN Heatmaps - frame k

Figure 4.18: Example of an accurate keypoints detection (a) leading to a large pose estimation error as a result of features switch (b).

inaccurate heatmaps detected at image frame k (b) lead to a large pose estimation despite a correct location of the 2D features at image frame k+1 (a). Ideally, such large heatmaps dispersions could be used to trigger a 2D-3D mismatch flag at step k+1. In this case, the SoftPosIT algorithm (David et al., 2004) could be used to solve for the pose, exploiting the fact that this algorithm assumes unknown feature correspondences. As SoftPosIT is an iterative solver, the estimated pose at frame k would be used during initialization.

Results for the selected scenario indicate that the estimated pose can be refined once the correspondences are handled by SoftPosIT. Specifically, an attitude error $E_R < 2^\circ$ can be achieved, proving that the proposed method could be used to refine the pose under wrong 2D-3D correspondences. Unfortunately, the validation of the proposed method over the entire image sequence of the V-bar trajectory showed multiple scenarios in which a large heatmaps dispersion does not correlate with a feature switch, leading to even worse accuracies after the iterative refinement. As such, different correlations shall be investigated to assess the robustness of this method. Nevertheless, the pose error analysis showed that the CNN can confuse similar features, when tested on a domain which is very different than the training one.

## 4.7. CHAPTER CONCLUSIONS

This Chapter introduced a novel on-ground validation framework to assess the performance of a monocular CNN-based pose estimation system on lab-generated space imagery. The performance of the proposed system is evaluated on a representative V-bar trajectory around a 1:25 mockup model of the Envisat spacecraft by recreating space-like illumination conditions and simultaneously operating two KUKA robots, in order to recreate the translational motion of the camera as well as the rotational motion of the target spacecraft. Thanks to the reconfigurability of the robotic arms, the proposed setup

is capable of recreating realistic rendezvous trajectories under multiple camera-target geometries. Moreover, the proposed calibration procedure guarantees reliable reference pose labels associated to each image of the generated trajectory, allowing the on-ground validation of the CNN pose estimation performance.

The domain shift problem typical of CNNs is tackled by introducing a novel data augmentation pipeline which includes both light and texture randomization. Results on the high exposure, 40° illumination Azimuth scenario indicate that over half of the V-bar trajectory is characterized by pose accuracies $E_T < 10\%$, $E_R < 10°$, an impressive result given the large domain gap between the synthetic training images and the GRALS-generated images. Specifically, these results highlight that texture randomization during training increases the CNN robustness against previously unseen target features, forcing the CNN to rely on the target shape instead of its textures. Moreover, preliminary analyses of the large pose estimation scenarios indicate that the adopted CNN undergoes feature switching when challenged with large domain shifts, suggesting that an iterative SoftPosIT refinement, triggered by monitoring the heatmaps dispersion pattern, could further improve the pose estimation accuracy.

# 5

# Adaptive CNN-based Relative Navigation

## 5.1. Introduction

Building on the pose estimation insights gained at a synthetic level, the on-ground validation carried out in Chapter 4 reported on the benefits of data augmentation and randomization to bridge the domain shift problem which characterizes synthetically-trained CNN when tested on realistic space imagery. The pose estimation analysis carried out on a representative V-bar scenario simulated at ESTEC's GRALS facility proved an increase in robustness towards HIL imagery: by testing the CNN on a laboratory mockup of Envisat characterized by previously unseen textures and illuminations, it was discovered that randomizing the texture of the target spacecraft during training allows the CNN to generalize textures and to focus on the shape of the target instead. This is expected to vary less from the synthetic rendering model to the realistic laboratory mockup. From a domain shift perspective, however, it was observed that the CNN detection accuracy considerably decreases under highly adverse illumination conditions or low camera exposures, suggesting that additional augmentation techniques are required to tackle the domain shift from an illumination standpoint (Pasqualetto Cassinis et al., 2022a).

In addition to the domain shift problem, additional challenges arise in relation to the applicability of CNNs for relative navigation in space. As already mentioned in Section 1.2, the interface between a CNN-based pose estimation method and a navigation filter has not been fully explored yet. Although the results obtained in Chapter 3 already suggested that a tightly-coupled approach should be more suited than a loosely-coupled one, limited results were provided at a navigation filter level on how to model the feature detection uncertainty into a representative measurements error covariance. In this context, the novel heatmaps-based method introduced in Chapter 3 has not been extensively tested on realistic space imagery. Specifically, it is still unclear whether the magnitude of

Parts of this chapter have been submitted in Pasqualetto Cassinis et al., 2022b.

such a covariance representation can effectively represent the feature detection uncertainty. Moreover, despite its promising results on ideal rendezvous scenarios with a high measurement frequency and synthetic images as measurements, the MEKF adopted in Section 3.6 could considerably decrease its performance if measurements from realistic imagery is acquired at lower frequencies. As such, it is important to extend the analyses carried out in Chapter 3 and address the navigation performance when low measurement frequencies occur. Finally and related to the CNN validation methodology, the online inference with the proposed HRNet model described in Chapters 3 and 4 was performed on a high-performance GPU. In order to address the applicability of the proposed CNN-based system in space, its performance should be validated on a representative space processor. In this context, processors with low-power consumption are currently being used to showcase the applicability of machine learning in space, including the Intel Movidius Myriad 2 and its successor Myriad X. The family of Myriad processors is capable of performing fast inferences while maintaining the power consumption well below 2 W (Giuffrida et al., 2021).



In the context of on-ground validation, the recreation of realistic rendezvous trajectories on-ground at ESTEC's GRALS facility has already been showcased in Chapter 4. In the GRALS setup, a V-bar trajectory around a mockup of the Envisat spacecraft was recreated by controlling both the target rotation and the translation of the monocular camera through two robotic arms, thus extending the testbed capability from the generation of static datasets to actual relative trajectories. However, only a static sun lamp was used to illuminate the target, leading to simplified illumination conditions. Furthermore, the calibration of the overall system was complicated by an inaccurate calibration of the target mockup, leading to lower accuracies in the reference pose labels compared to Stanford's robotic Testbed for Rendezvous and Optical Navigation (TRON). In a recent effort to solve the above issues, the capabilities of the TRON facility were extended towards the generation of the Satellite Hardware-In-the-loop Rendezvous Trajectories (SHIRT) dataset, a dataset of synthetic and realistic imagery of two close-proximity trajectories around the Tango spacecraft with high-fidelity calibration and more realistic illumination conditions of a typical LEO orbit (Park and D'Amico, 2022a). As a result, new possibilities opened up in the framework of on-ground validation of pose estimation systems in rendezvous scenarios typical of ADR and OOS missions.

Building on the findings detailed in Chapters 3-4, this Chapter aims at solving the above mentioned challenges in a sequential fashion. First, a data augmentation pipeline centered on light augmentation is introduced to extend a texture-based data augmentation and solve the domain shift problem. An existing technique, introduced by Sakkos et al. (2019) for terrestrial applications, is exploited to generalize the illumination conditions during the CNN training. The performance of the CNN-based pose estimation system is then evaluated on realistic imagery of the SPEED+ dataset at a pose estimation level, by validating the system on both a GPU and on the Myriad X processor. Next, the CNN system is combined with an Unscented Kalman Filter (UKF) to address the performance of the proposed system at a navigation filter level. To this end, the SHIRT dataset is used in the evaluation in order to compare the filter performance in synthetic scenarios with the

performance on realistic lab imagery. To cope with the challenges in the representation of the measurements uncertainty, an adaptive scheme is proposed in which the measurement error covariance is estimated online by scaling the heatmaps-based representation based on filter innovation.

In summary, the main contributions of this Chapter are:

1. To propose a novel data augmentation scheme based on light randomization in order to improve the CNN performance on realistic space imagery.
2. To verify the implementability and performance of the adopted CNN on a representative space processor.
3. To propose a covariance adaptation method based on CNN heatmaps to capture measurements uncertainty within the navigation filter.
4. To validate a tightly-coupled, CNN-based UKF on realistic trajectories generated on-ground using the TRON facility.



The Chapter is organized as follows. Section 5.2 introduces the proposed validation framework. The TRON testbed and the image acquisition procedure are described in Section 5.3. In Section 5.4, the CNN training, validation and testing phases are detailed, with special focus to the data augmentation pipeline. Section 5.5 presents the preliminary pose estimation results on SPEED+ test datasets. Next, the proposed navigation filter is described in Section 5.6 together with the novel adaptive scheme for the measurement error covariance. Sections 5.7-5.8 describe the simulation environment and show the navigation results on the two recreated close-proximity trajectories. Finally, Section 5.9 provides the main conclusions and recommendations.

## 5.2. VALIDATION FRAMEWORK

The validation pipeline of the proposed navigation system is shown in Figure 5.1 and consists of the following main stages:

1. **Calibration Procedure and Image Acquisition**: The adopted monocular camera is calibrated with respect to a scaled 1:2 mockup model of the Tango spacecraft via a dedicated procedure, in order to associate accurate reference labels of the pose between the camera and the mockup. Once the setup is calibrated, laboratory images of the target mockup are generated by the camera and associated to their corresponding pose labels. Trajectory design and execution is carried out in order to create a representative laboratory database (SPEED+) as well as to perform rendezvous trajectories around the mockup (SHIRT). The calibration procedure and the trajectory generation pipeline are taken from Park, Martens, et al. (2021) and Park and D'Amico (2022a), respectively.
2. **Dataset Augmentation and CNN Training**: A keypoints-based CNN is trained and validated on augmented SPEED+ datasets. The augmentation is performed by

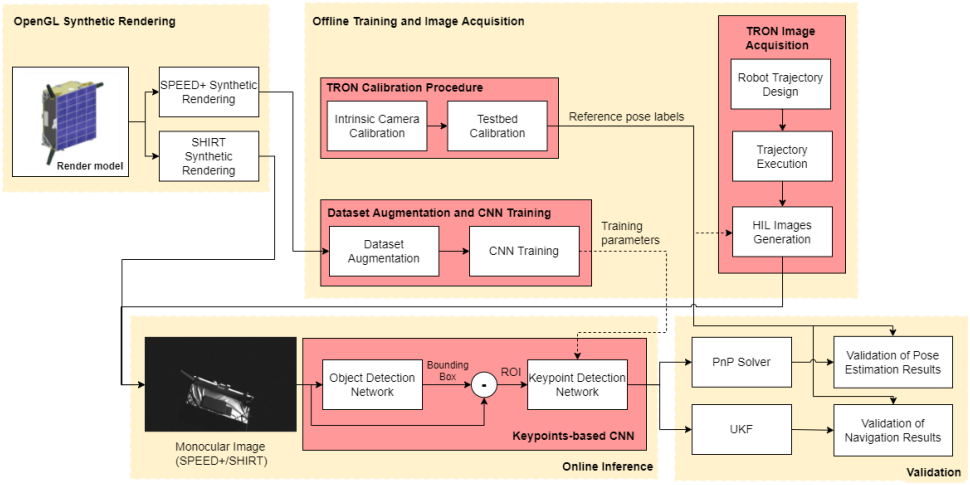



Figure 5.1: Illustration of the proposed on-ground validation of the CNN-based pose estimation system.

5

introducing image noise, random background and random illuminations into the synthetically-generated images of SPEED+. The synthetic images are created using the OpenGL-based graphics renderer.

3. **Online Inference**: The keypoints-based CNN is tested on both synthetic and HIL test images. For the SPEED+ images, the pose is estimated by feeding a PnP solver with the detected keypoints as well as with the intrinsic camera parameters and the 3D model of the Tango spacecraft. Conversely, for the SHIRT images the detected keypoints are fed into a CNN-based UKF to additionally estimate the relative translational and rotational velocities.

4. **Post-Processing and Validation**: The results of the proposed CNN-based system on the HIL images are validated against the reference pose labels, derived from the calibration setup. This is performed at both pose estimation and navigation levels.

### 5.2.1. RELATIVE NAVIGATION

This work considers a servicer spacecraft flying relative to a target spacecraft in a Low Earth Orbit (LEO), with the relative motion being described in the LVLH reference frame co-moving with the servicer (Figure 2.7). Furthermore, it is assumed that the servicer is equipped with a single monocular camera. The relative attitude of the target with respect to the servicer can then be defined as the rotation of the target body-fixed frame $\mathcal{B}$ with respect to the servicer camera frame $\mathcal{C}$, where these frames are fixed to each spacecraft's body. The vector from the origin of the camera frame to the origin of the target frame defines their relative position. Together, these two quantities characterize the pose. This information can then be transformed from the camera frame to the servicer's center of mass by accounting for the pose of the camera with respect to the LVLH frame. Beside the Cartesian representation of the relative motion between the servicer and target spacecraft, the relative state can also be parametrized as a function of the absolute orbital

elements of the two spacecraft. This Chapter uses the Relative Orbital Elements (ROE) state introduced by D'Amico (2010), which are defined in terms of the classical Keplerian orbital elements as:

$$\boldsymbol{\delta\alpha} = \begin{bmatrix} \delta a \\ \delta\lambda \\ \delta e_x \\ \delta e_y \\ \delta i_x \\ \delta i_y \end{bmatrix} = \begin{bmatrix} (a_t - a_s)/a_s \\ (M_t - M_s) + (\omega_t - \omega_s) + c_{i_s}(\Omega_t - \Omega_s) \\ e_t c_{\omega_t} - e_s c_{\omega_s} \\ e_t s_{\omega_t} - e_s s_{\omega_s} \\ i_t - i_s \\ s_{i_s}(\Omega_t - \Omega_s) \end{bmatrix} \tag{5.1}$$

where the subscripts $t$ and $s$ indicate the target and servicer spacecraft respectively, and $s_i$ and $c_w$ represent the sine and cosine of the argument of perigee $\omega$ and inclination $i$, respectively. Notice that this set of ROEs is nonsingular for circular orbits, but is singular for equatorial orbits.

## 5.3. TRON TESTBED

The TRON facility at SLAB, visualized in Figure 5.2, includes a control room and an 8 × 3 × 3 m simulation room which consists of various components and machineries to (i) simulate the vision-based rendezvous trajectory of a servicer spacecraft with a camera to a target spacecraft, and (ii) emulate the high-fidelity spaceborne illumination conditions to maximize the realism of the images captured by the camera. TRON comprises two 6 degrees-of-freedom KUKA robot arms and a set of Vicon motion track cameras to reconfigure an arbitrary pose between a camera and a target mockup model, as well as multiple Earth albedo light boxes and a sun lamp to simulate high-fidelity spaceborne illumination conditions. The calibration of the facility is performed via a dedicated multi-source Robot/World Hand/Eye (RWHE) calibration procedure which fuses readings from KUKA and Vicon to accurately estimate the pose of the adopted monocular camera with respect to the target mockup. Millimeter-level position accuracy and millidegree-level orientation accuracy were obtained on a subset of close-range images of the Tango mockup. The reader is referred to Park, Bosse, et al. (2021) for a detailed description.

### 5.3.1. SPEED+ DATASET

The SPEED+ dataset was created in the TRON facility to be used in the ongoing international Satellite Pose Estimation Challenge[1], with the main objective to evaluate and compare the robustness of machine learning models trained on synthetic images (Park, Martens, et al., 2021). The dataset is built upon the existing SPEED dataset (Kisantal et al., 2020) by increasing the number of synthetic and HIL images, whilst extending the illumination conditions simulated in the facility and improving the accuracy of the pose labels. SPEED+ consists of 59,960 synthetic images generated in OpenGL and 6,740 - 2,791 realistic images generated in TRON with light boxes and sun lamp, respectively (see Figure 5.3). Figure 5.4 illustrates sample SPEED+ images for the synthetic, `lightbox`, and `sunlamp` domains. Overall, the test images of SPEED+ extend the full orientation space

[1] https://kelvins.esa.int/pose-estimation-2021/challenge/

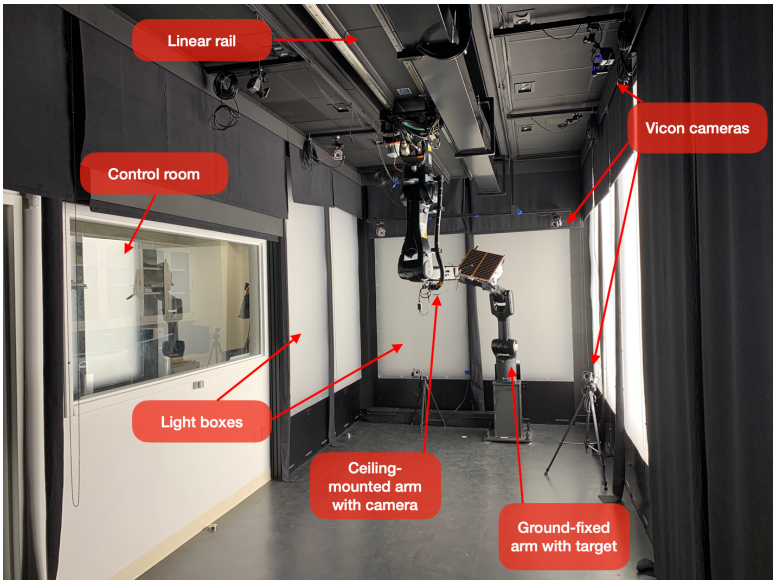



Figure 5.2: TRON simulation room and its components (Park, Bosse, et al., 2021).

and distance up to 10 m with realistic re-creation of Earth albedo and direct sunlight present in spaceborne imagery.

### 5.3.2. SHIRT DATASET

A detailed overview of the generation of the SHIRT dataset can be found in Park and D'Amico (2022a). For the HIL images, the first step is to design a rendezvous trajectory which resembles a typical close-proximity scenario. Next, the motion of the two KUKA robots is commanded in order to recreate the desired trajectory, taking into account the scale of the target mockup and the constraints of the facility. In order to simulate representative illumination conditions, the *true* location of the Sun is used to intermittently switch the light boxes on/off and capture the correct inclination of the Sun with respect to the mockup. Figure 5.5 shows a comparison of synthetic and TRON images for the same poses, where the synthetic images were generated for the same illumination and trajectory inputs. As can be seen, the TRON illumination manages to capture the correct inclination of the Sun used in the synthetic renderings. Besides, a clear domain gap is present between synthetic and TRON images, ensuring a the validation of the proposed CNN-based system in challenging domain shifts.

## 5.4. CONVOLUTIONAL NEURAL NETWORK

Following the promising results reported in Chapters 3 and 4, the HRNet architecture (Sun et al., 2019) is chosen at an image processing level to extract 12 pre-defined keypoint features of the Tango spacecraft from an input 2D image. The corners of the main spacecraft body and the extremities of its three antennas are pre-selected as keypoint


Figure 5.3: Visualization of the TRON facility simulation room layout for the generation of the SPEED+ images (Left). Ten light boxes (L1 - L10) and three positions of the sun lamp (S1 - S3) are marked and noted. The half-scale Tango spacecraft mockup model is illuminated by two light boxes L1,L2 (Middle) or by the sun lamp placed at S1 (Right) (Park, Martens, et al., 2021).



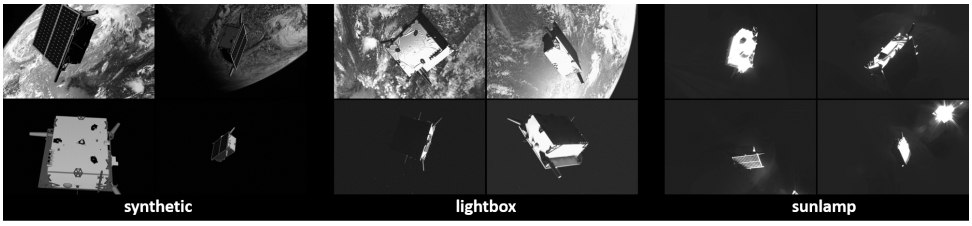



Figure 5.4: Example images from different domains of SPEED+ (Park, Martens, et al., 2021)

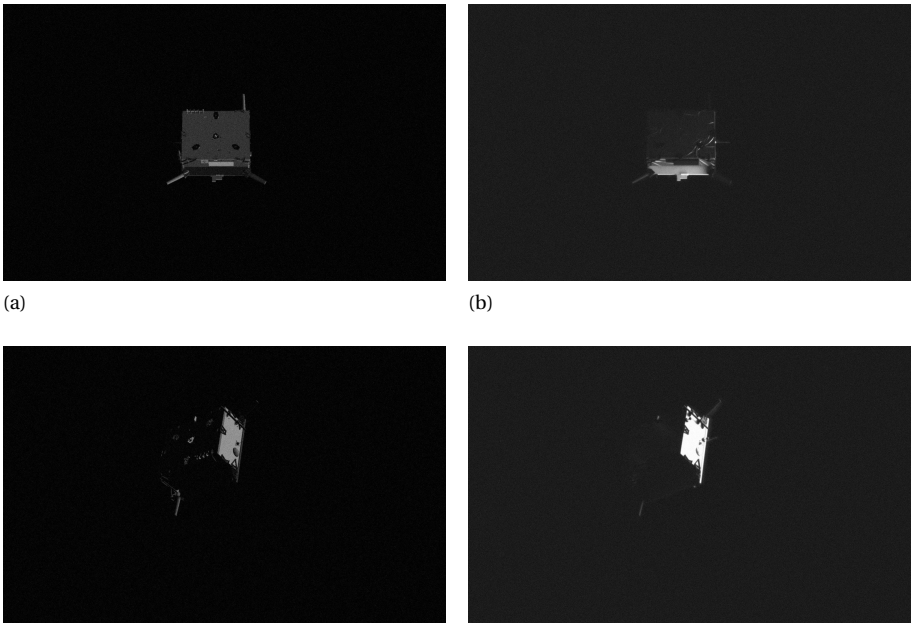
features and used to train the CNN. These features are chosen since they already proved reliable in recent CNN trainings performed on the Tango spacecraft (Chen et al., 2019; Kisantal et al., 2020).

### 5.4.1. DATA AUGMENTATION PIPELINE

In Figure 5.6, the first step of the proposed pipeline for the datasets augmentation and randomization consists in taking the ideal synthetic images of the Tango spacecraft from the SPEED+ dataset, which already includes images with the Earth in the background. Similar to the data augmentation described in Chapter 4, a noise pipeline is then applied in order to augment the training and validation datasets with the following noise models:

1. Gaussian, shot, impulse and speckle noise.
2. Gaussian, defocus, motion and zoom blurs.
3. Spatter, color jitter and random erase.

Finally, light augmentation is introduced to generalize the illumination conditions during training. The proposed approach uses the data augmentation method introduced by



(a) (b)

(c) (d)

Figure 5.5: Synthetic (a and c) and HIL (b and d) sample images for two representative poses.

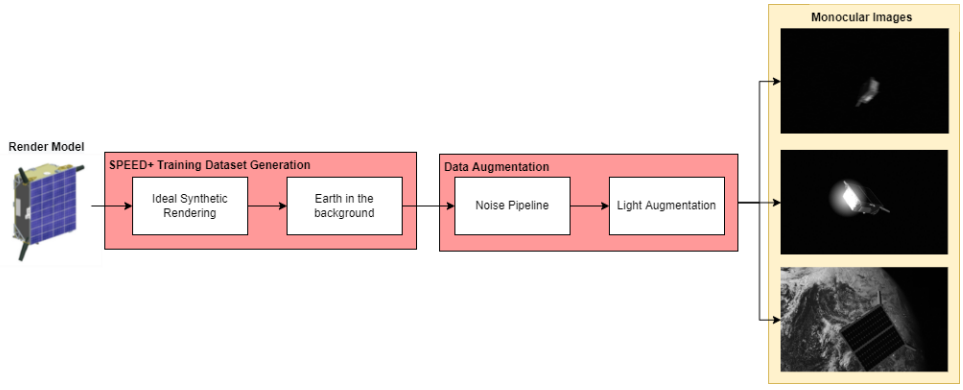



Figure 5.6: SPEED+ Dataset Augmentation Pipeline.

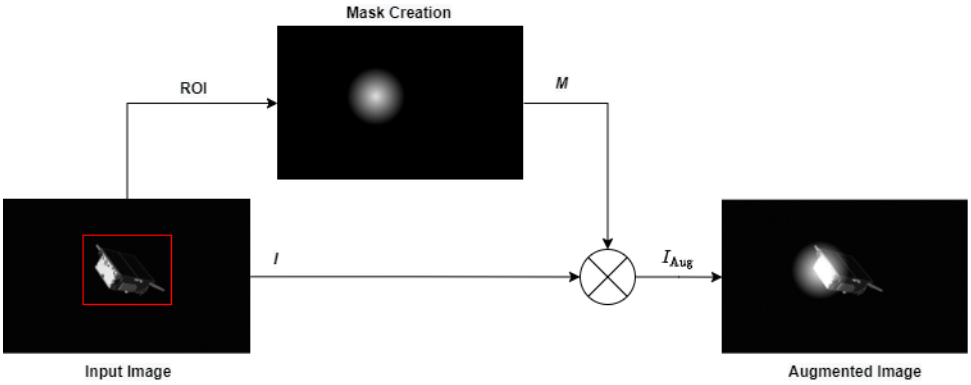



Figure 5.7: Description of light augmentation.

Sakkos et al. (2019) for terrestrial applications and extends it to the SPEED+ dataset. Figure 5.8 illustrates the light augmentation pipeline for a sample image $I$ of the SPEED+ dataset. First, a ROI is extracted around the target spacecraft and used to create an illumination mask $M$ at random locations around the target. The mask consists of an illumination circle with centre $p = I(w,h), w \in W, h \in H$ and diameter $d = k \times \min(W,H)$, where $H, W$ are the height and width of the extracted ROI and $k \in (\frac{1}{5}, \frac{1}{2})$. Since uniformly modifying all pixels within the illumination circle generates unrealistic results, the Euclidean Distance Transform (EDT) is further applied to the mask in order to model light attenuation (Sakkos et al., 2019). For each pixel in the circle, the EDT assigns a number that is the distance between that pixel and the nearest nonzero pixel.

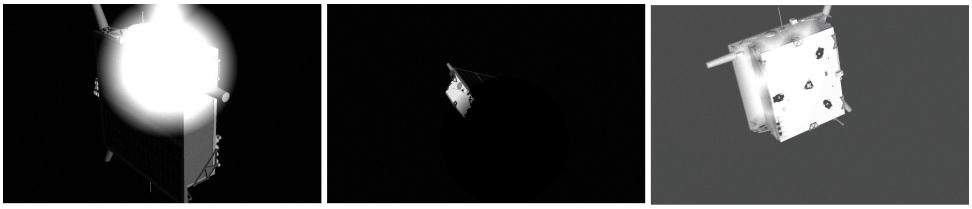
Once the mask is created, the original image is augmented to create three distinct effects:

$$\boldsymbol{I}_{\text{Aug}} = \begin{cases} \boldsymbol{I} + \boldsymbol{M}z_1 & \text{Bright filter} \\ \boldsymbol{I} - \boldsymbol{M}z_1 & \text{Dark filter} \\ \boldsymbol{I} + \boldsymbol{M}z_1 + z_2 & \text{Local and global filters} \end{cases} \tag{5.2}$$

where $\boldsymbol{I}_{\text{Aug}}$ is the augmented image, $\boldsymbol{M}$ is a mask of the same size as the input image $\boldsymbol{I}$, and $z_1, z_2$ are random integers. Figure 5.8 illustrates the three filters applied to sample image $I$. Notably, several illumination effects can be introduced in the datasets by interchanging these filters whilst varying the disc diameter and location. In the current implementation, 50% of the synthetic images are augmented with light augmentation, of which 50% have a dark filter, 25% a bright filter, and 25% local and global filters.

### 5.4.2. TRAIN, VALIDATION AND TEST

Table 5.1 lists the number of images used in the Train, Validation, and Test datasets together with the data augmentation breakdown. The ideal synthetic images of the SPEED+ dataset are first split into 80:20 train/validation sets. Next, the augmentation pipeline described in Section 5.4.1 is used to extend both sets. During training, the validation dataset is used beside the training dataset to compute the validation losses



*Bright filter* *Dark filter* *Local and global filters*

Figure 5.8: Example of different light augmentation effects.

Table 5.1: Description of Train, Validation and Test datasets together with data augmentation breakdown.

| **Dataset** | **Synthetic** | `lightbox` | `sunlamp` |
|---|---|---|---|
| Train | 47,966<br>(80% Noise Pipeline)<br>(50% Light Augmentation) | - | - |
| Validation | 11,994<br>(80% Noise Pipeline)<br>(50% Light Augmentation) | - | - |
| Test | - | 6,740 | 2,791 |



and avoid overfitting. The Adam optimizer (Kingma and Ba, 2015) is used with a cosine decaying learning rate with initial value of $10^{-3}$ and decaying factor of 0.1. Finally, the CNN is tested on the `lightbox` and `sunlamp` sets. In this way, the performance of a CNN trained solely on synthetic imagery can be assessed on the realistic imagery simulated in TRON. A Tesla P100-PCIE-16GB GPU is used for both training and testing.

### 5.4.3. MYRIAD X IMPLEMENTATION

Next to the GPU implementation, the adopted HRNet pipeline is tested on the Myriad X processor shown in Figure 5.9. Inference is executed on two deep neural Vector Processing Units (VPU). Notably, the inference can be parallelized by allocating some tasks on the Neural Compute Engine and others on one of the sixteen Streaming Hybrid Architecture Vector Engine (SHAVE) cores. Due to its low thermal budget ($\approx$ 1.5 W), Myriad X is ideally suited for in-orbit applications. The model conversion from Python's PyTorch machine learning framework to Myriad X is described in Figure 5.10. First of all, the GPU-trained model is exported as .onnx file and optimized using the OpenVino model optimizer[2]. This optimizer is a cross-platform command-line tool that facilitates the transition between training and deployment environments. It performs static model analysis and adjusts deep learning models for optimal execution on end-point target devices. Then, the .xml file generated during the model optimization is used by a dedicated tool to compile a

[2] https://docs.openvino.ai/latest/index.html

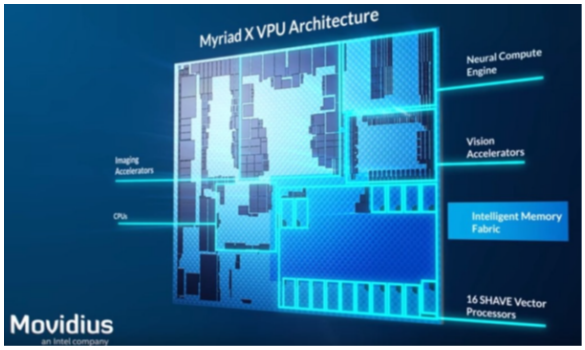



Figure 5.9: Illustration of the Myriad X architecture.

Myriad-compatible model. Finally, the Ubotica CVAI Toolkit software [3] allows for the deployment of applications involving image processing and inference.



## 5.5. POSE ESTIMATION

Following the promising pose estimation results described in Sections 3.7.1 and 4.6, the Efficient Perspective-n-Points (EPnP) method followed by Gauss-Newton refinement (Lepetit et al., 2009) is selected to estimate the pose from a set of detected features. The error metric introduced in Section 3.7 and Section 4.6 is used, namely:

$$E_T = \left\| \boldsymbol{t}^C - \hat{\boldsymbol{t}}^C \right\| \tag{5.3}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_s & \boldsymbol{\beta_v} \end{bmatrix} = \boldsymbol{q} \otimes \hat{\boldsymbol{q}} \tag{5.4}$$

$$E_R = 2 \arccos\left(|\beta_s|\right). \tag{5.5}$$

where the translational error is expressed as the norm of the difference between the estimated relative position $\hat{\boldsymbol{t}}^C$ and the ground truth $\boldsymbol{t}_C$, and the rotational error is expressed in terms of the Euler axis-angle error between the estimated quaternion $\hat{\boldsymbol{q}}$ and the ground truth $\boldsymbol{q}$. Furthermore, a combined score is created by combining both position and attitude errors,

$$E_{\text{pose}} = E_R + \frac{E_T}{\left\| \boldsymbol{t}^C \right\|}. \tag{5.6}$$

Note that when evaluated on SPEED+ `lightbox` and `sunlamp` samples, a modified SPEED metric $E^*_{\text{pose}}$ (Park, Martens, et al., 2021) is used to zero out the errors smaller than the thresholds based on the TRON calibration, i.e. for individual sample,

$$E^*_{\text{pose}} = \begin{cases} 0 & \text{if } E_R < 0.169^\circ \quad \text{and} \quad \frac{E_T}{\left\| \boldsymbol{t}^C \right\|} < 2.173 \quad \text{mm/m} \\ E_{\text{pose}} & \text{otherwise.} \end{cases} \tag{5.7}$$

[3] https://ubotica.com/product/specifications/

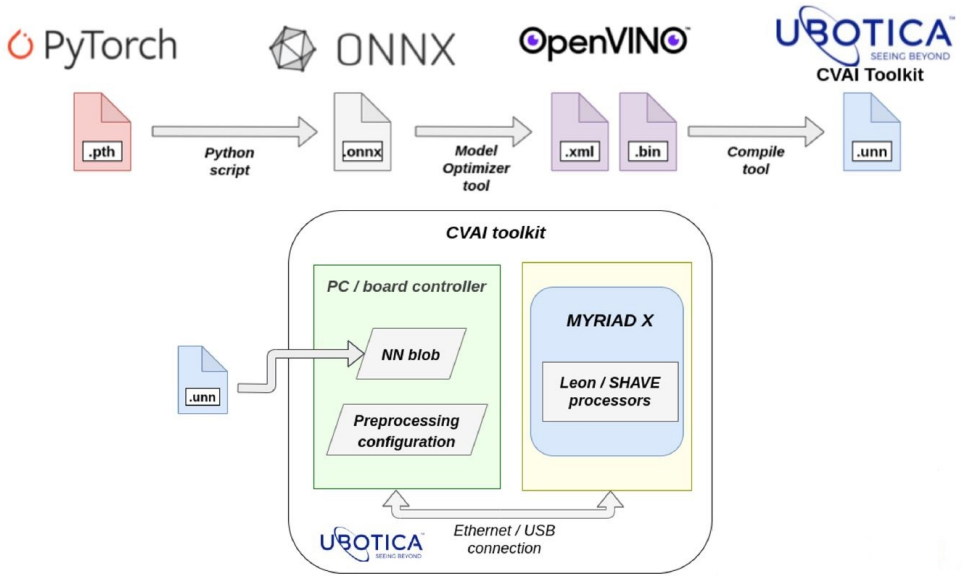


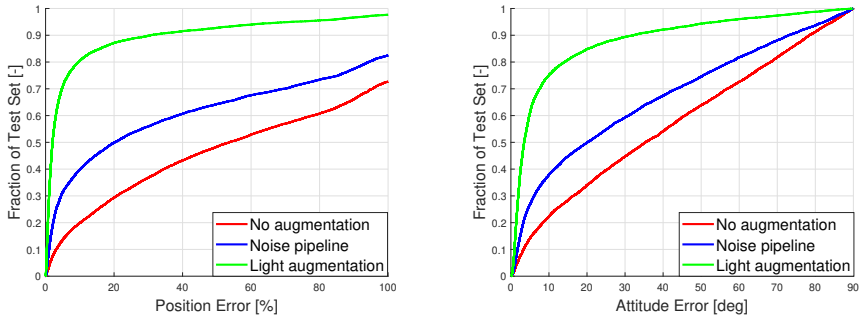

Figure 5.10: Illustration of the functional flow for the model conversion to Myriad X.

### 5.5.1. POSE ESTIMATION RESULTS

The pose estimation results of the CNN-based system in Figure 5.1 are presented for the SPEED+ test dataset, in order to evaluate the capability of the CNN to bridge the domain gap between the synthetic training and the TRON imagery. Additionally, the proposed pose estimation system is evaluated on the PRISMA25 dataset, which consists of 25 flight images of the Tango spacecraft from the PRISMA mission (D'Amico et al., 2013). Despite the limited number of images in PRISMA25, the comparative study between SPEED+ and actual flight images allows an assessment of the scalability of the proposed data augmentation method to different target domains, whilst providing insight to the applicability of HIL images as a surrogate of flight images for validation. The pose estimation metrics introduced in Section 3.7 are used.

#### SPEED+

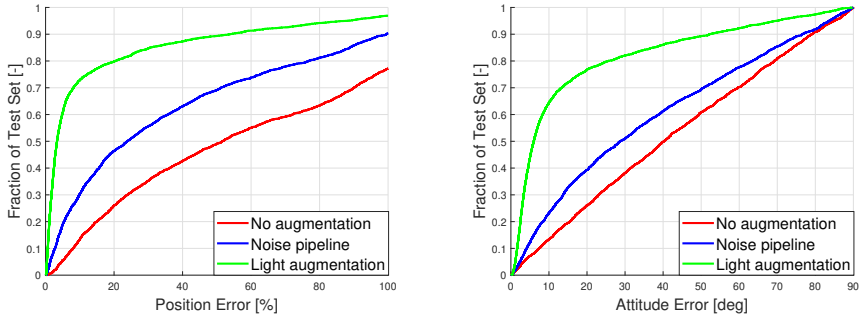
Figures 5.11-5.12 show the pose estimation results in terms of the Cumulative Distribution Function (CDF) across the test datasets. Referring to Section 5.4.1, the results are reported for an augmentation-free CNN training, a training with the noise pipeline, and a training with both noise and light augmentation. This is done in order to assess the impact of each augmentation on the CNN performance. As can be seen, the introduction of light augmentation into the CNN training greatly improves the overall pose estimation performance on both the `lightbox` and the `sunlamp` subsets. Specifically, pose errors $E_T < 0.1 \left\| \boldsymbol{t}^C \right\|$, $E_R < 10^\circ$ and $E_T < 0.1 \left\| \boldsymbol{t}^C \right\|$, $E_R < 13^\circ$ are achieved in 80% of the `lightbox` and `sunlamp` subsets, respectively. Figure 5.13 illustrates highly accurate pose estimation results on a representative subset of TRON images with lightbox illuminations, Earth in the background, and sunlamp illuminations. Conversely, Figure 5.14 illustrates two




(a) (b)

Figure 5.11: Cumulative Distribution Function of Position (a) and Attitude (b) errors over the `lightbox` subset of SPEED+.






(a) (b)

Figure 5.12: Cumulative Distribution Function of Position (a) and Attitude (b) errors over the `sunlamp` subset of SPEED+.

representative scenarios characterized by large pose errors. As can be seen, near-eclipse illumination conditions and highly adverse sunlamp reflections can still jeopardize the CNN performance, despite the adopted data augmentation pipeline. Notably, similar effects can be observed in some of the highly accurate pose estimates (Figure 5.13), suggesting that the CNN performance could be affected by small visual artifacts not visible by the human-eye.

### PRISMA25

Figure 5.15 shows the CDF pose estimation results for the PRISMA25 dataset in terms of the total SPEED score. The results are compared with the `lightbox` and `sunlamp` scenarios in order to assess the scalability of the proposed light augmentation on flight images. Notably, 60% of the images are characterized by highly accurate poses with mean errors $E_T = 0.5$ m, $E_R = 2.5°$, demonstrating the effectiveness of the proposed augmenta-

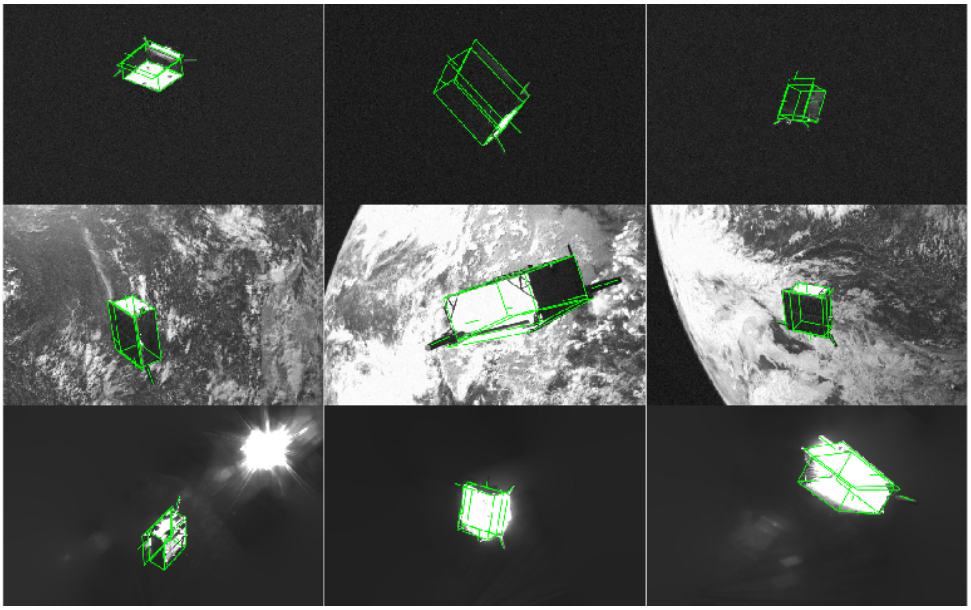


Figure 5.13: Example of pose estimation results for subsets of the `lightbox` (Top), `lightbox` with Earth in the background (Middle) and `sunlamp` (Bottom) images. The wireframe model of the Tango spacecraft is projected based on the estimated pose.

tion pipeline. Notice also that the CDF of PRISMA25 is bounded by the `lightbox` and `sunlamp` CDFs for most of the scores, suggesting that the TRON illuminations represent realistic best/worst case bounds for illuminations in actual space imagery.

Besides, Table 5.2 lists the mean errors results associated to the CDF in Figure 5.15, whilst comparing the proposed light augmentation method with another augmentation pipeline which exploits texture randomization (Jackson et al., 2018). As can be seen, light augmentation provides a considerable improvement over texture randomization.

### MYRIAD X

Similar to the analysis on the Envisat scenario described in Section 4.6 and in order to ease the comparison between the GPU implementation and the Myriad X implementation, the pose estimation error is categorized into high, medium, and low accuracies:

- High accuracy: $E_T < 5\%$, $E_R < 2°$,
- High/medium (Medium) accuracy: $E_T < 10\%$, $E_R < 5°$,
- High/medium/low (Low) accuracy: $E_T < 10\%$, $E_R < 10°$.

Table 5.3 lists the pose estimation results together with the mean inference time, herewith defined as the time needed to load the input image and detect the heatmaps. Notably, this inference time does not account for post-processing tasks, such as the extraction of keypoints' location from their heatmaps. As can be seen, the HRNet performance on

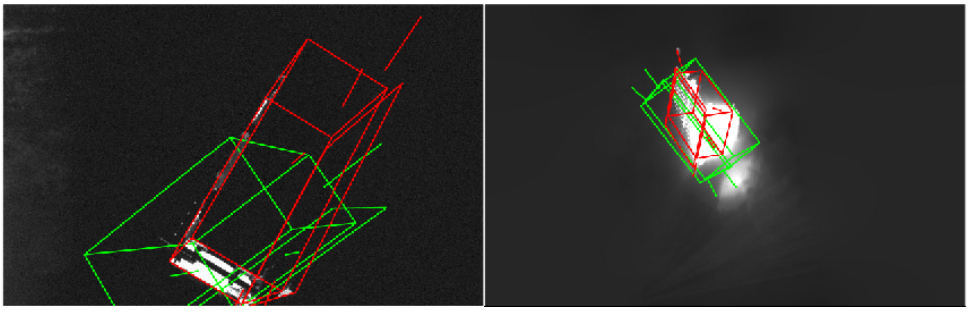


Figure 5.14: Example of inaccurate pose estimation results due to near-eclipse illuminations (Left) and adverse sunlamp reflections (Right). The wireframe model of the Tango spacecraft is projected based on the true (red) and estimated (green) poses.



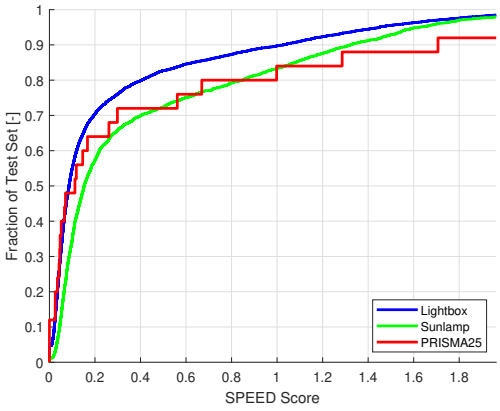
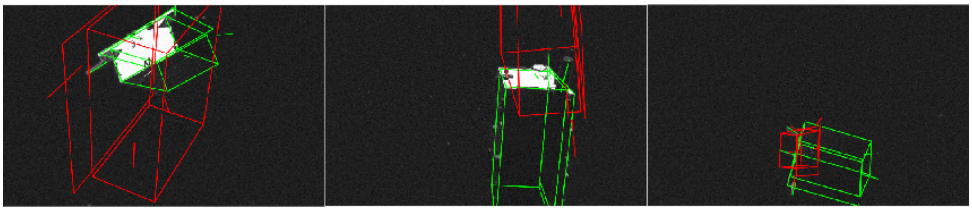


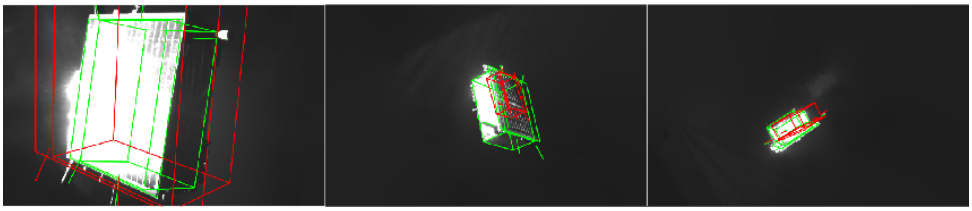

Figure 5.15: SPEED score Cumulative distribution function in PRISMA25 compared to `lightbox` and `sunlamp`.

Table 5.2: Pose estimation results on PRISMA25 dataset.

| CNN Model | Data Augmentation | SPEED Score | $E_T$ [m] | $E_R$ [deg] |
|---|---|---|---|---|
| KRN (Park, Martens, et al., 2021) | None | 1.98 | 7.02 | 77.5 |
| | Texture Rand. | 1.43 | 4.03 | 59.7 |
| HRNet | None | 1.76 | 11.17 | 42.1 |
| | Image Noise | 0.96 | 4.22 | 32.1 |
| | **Image Noise/Light Rand.** | **0.43** | **2.1** | **13.5** |



(a) `lightbox`



(b) `sunlamp`



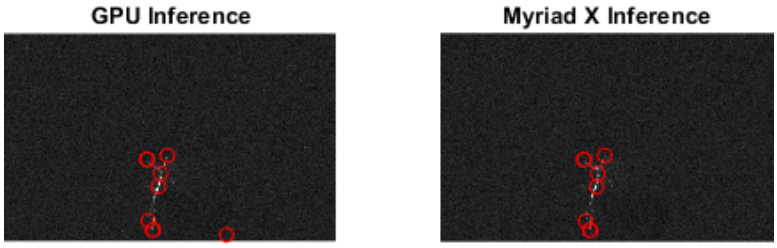
Figure 5.16: The wireframe model of the Tango spacecraft is projected based on the GPU (green) and Myriad X (red) inferences for representative images of the `lightbox` (Top) and `sunlamp` (Bottom) datasets. As can be seen, near-eclipse conditions and challenging reflections on the target could lead to a lower performance of HRNet on Myriad X.

Myriad X compares well with the performance on the GPU, with the same percentage of high, medium, and low pose accuracies on both the `lightbox` and the `sunlamp` datasets. Moreover, the mean inference time on Myriad X across both datasets is only around x2 slower than on the GPU. Considering that a Myriad X consumes only 1.5 W on average, as opposed to the 14 W average on the GPU, these results showcase a x6 better inference per Watt on Myriad X. This aspect becomes relevant considering that multiple Myriad can fly together and parallelize the HRNet tasks during feature detection.

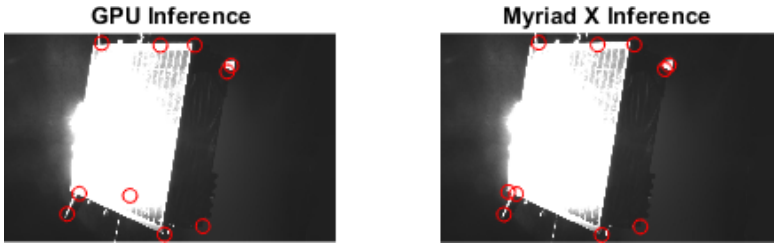
Table 5.3: Performance results of GPU and Myriad X on the `lightbox` and `sunlamp` datasets.

| | Mean Inference Time [ms] | High Accuracy | Medium Accuracy | Low Accuracy |
|---|---|---|---|---|
| | | **`lightbox`** | | |
| GPU | 67.9 | 30% | 59% | 71% |
| Myriad X | 135.6 | 29% | 59% | 71% |
| | | **`sunlamp`** | | |
| GPU | 75.7 | 12% | 40% | 58% |
| Myriad X | 138.6 | 12% | 40% | 58% |

Despite a comparable performance, however, a few scenarios were identified in which the pose accuracy greatly decreases on Myriad X. Figure 5.16 shows a subset of these scenarios for both the `lightbox` (Top) and `sunlamp` (Bottom) datasets. Interestingly, either the near-eclipse conditions or the strong reflections on the target body are suspected to be the main challenges for the HRNet implementation on Myriad X. This suggests that




(a) `lightbox`




(b) `sunlamp`

Figure 5.17: Keypoint detection results on GPU and Myriad X for two representative SPEED+ images. Due to the extreme near-eclipse conditions in the `lightbox` image (Top), the HRNet inference on MyriadX fails at detecting the keypoint located in bottom-right corner of Tango. Likewise, due to the extreme reflections in the `sunlamp` image (Bottom), the HRNet inference on Myriad X fails at detecting the occluded back corner.

the HRNet model conversion could decrease the complexity of the network and, in turn, lead to a worse keypoint detection performance, when large domain shifts occur. To investigate this aspect further, the keypoint detections associated to two of the scenarios shown in Figure 5.16 are analyzed in Figure 5.17. For the `lightbox` dataset, the third and most challenging scenario is selected, in which the Tango spacecraft is hardly visible due to the adverse Sun-camera-target geometry. As can be seen, the GPU inference manages to locate a corner on the poorly visible side of the target object, leading to a reliable set of features handled by the P$n$P solver despite the inaccuracies of some of the visible features. Conversely, the Myriad X inference can only locate the features on the visible side of the target object. As a result, the inaccuracies of the visible features cannot be compensated by the additional corner, leading to a large pose error. Similarly, the keypoint detection comparison on the selected `sunlamp` image shows that the Myriad X inference cannot detect an occluded corner which is detected by the GPU implementation. Although these observations are made on a limited set of images, they suggest that the model conversion to Myriad X could potentially jeopardize the pose estimation accuracy in highly adverse illumination conditions or large domain shifts from the training set.

## 5.6. NAVIGATION FILTER

Despite its promising results on ideal rendezvous scenarios with a high measurement frequency and synthetic images as measurements, the MEKF adopted in Section 3.6 could considerably decrease its performance under low measurement frequencies and

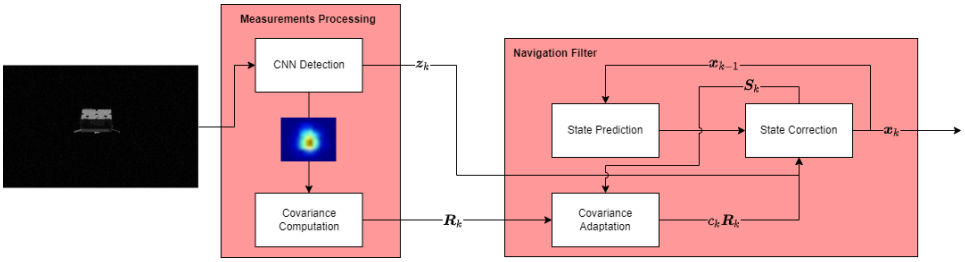



Figure 5.18: High-level description of the UKF flow together with the measurements processing and adaptive covariance. At step $k$, the CNN measurements $\boldsymbol{z}_k$ and the measurement error covariance $\boldsymbol{R}_k$ are used to correct the prediction of the filter state $\boldsymbol{x}_k$. The measurement error covariance is adapted with the filter innovation $\boldsymbol{S}_k$ by estimating a scaling coefficient $c_k$.

realistic space imagery, due to its linearization of both the relative state and the measurement equation (Eqs. 3.19-3.33). For these reasons, a UKF (Julier and Uhlmann, 2004) is used instead in this work, in order to capture the system nonlinearity via the unscented transform, which uses a finite set of deterministic samples instead of linearizing the nonlinearity itself. Specifically, the Unscented Quaternion Estimator (USQUE) introduced by Crassidis and Markley (2012) is adopted. The USQUE has proven more robust under low measurement acquisition frequency when compared to a more standard EKF/MEKF. Furthermore, the expected error is generally lower than the EKF, and the filter iterations avoid the derivation of Jacobian matrices (Crassidis and Markley, 2012). Similar to the MEKF, a three-component attitude-error vector is used to represent the quaternion error vector.

In a standard EKF, the state vector for pose estimation based on ROE is a 13-dimension vector composed of the relative ROE as well as the relative quaternion and rotational velocity,

$$\boldsymbol{x} = \begin{bmatrix} a_s \boldsymbol{\delta\alpha}^T & \boldsymbol{q}_B^{C\,T} & \boldsymbol{\omega}_{B/C}^{C\,T} \end{bmatrix}^T, \tag{5.8}$$

where $a_s$ represents the semi-major axis of the servicer spacecraft, $\boldsymbol{q}_B^C = [q_0 \quad \boldsymbol{q_v}]$ is the quaternion set that represents the relative attitude, and $\boldsymbol{\omega}_{B/C}^C$ is the angular velocity of the target with respect to the camera, expressed in the camera frame. In the USQUE, the modified state vector propagated inside the filter becomes a 12-dimension vector,

$$\tilde{\boldsymbol{x}} = \begin{bmatrix} a_s \boldsymbol{\delta\alpha}^T & \boldsymbol{\delta p}^T & \boldsymbol{\omega}_{B/C}^{C\,T} \end{bmatrix}^T, \tag{5.9}$$

where $\boldsymbol{\delta p}$ is four times the Modified Rodrigues Parameters (MRP) $\boldsymbol{\sigma}$,

$$\boldsymbol{\delta p} = 4\boldsymbol{\sigma} = 4\frac{\boldsymbol{q_v}}{1+q_0}. \tag{5.10}$$

A high-level description of the proposed navigation pipeline is provided in Figure 5.18. Each subsystem is described in the following sections.

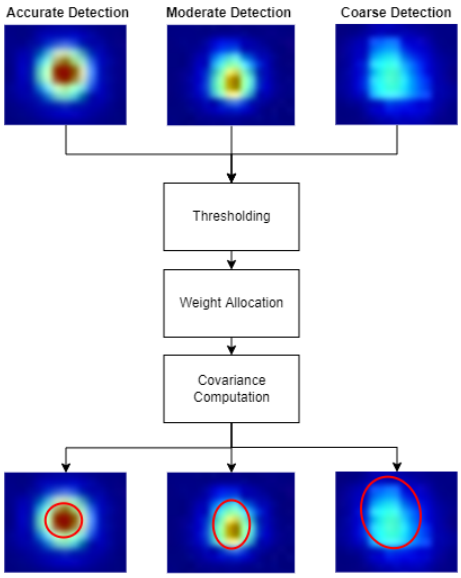



Figure 5.19: Schematic of the procedure followed to derive covariance matrices from CNN heatmaps. Accurate, moderate and coarse detections are represented. The displayed ellipses are derived from the computed covariances by assuming a confidence interval $\mathbf{1\sigma = 0.68}$.

### 5.6.1. MEASUREMENT ERROR COVARIANCE COMPUTATION

The measurement error covariance is associated to each feature directly from the heatmaps detected by the CNN, rather than from the computation of the image gradient around each feature. Referring to Section 3.4, this method consists in extracting the $i$th non-zero pixel around the heatmap's peak and derive a covariance matrix $\boldsymbol{C}_i$,

$$\boldsymbol{C}_i = \begin{bmatrix} \text{cov}(x,x) & \text{cov}(x,y) \\ \text{cov}(y,x) & \text{cov}(y,y) \end{bmatrix}, \tag{5.11}$$

where

$$\text{cov}(x,y) = \sum_{i=1}^{n} w_i (x_i - p_x) \cdot (y_i - p_y). \tag{5.12}$$

Here, $w_i$ is a normalized weight based on the gray intensity $I_i$ at each pixel location, and $n$ is the number of pixels in each feature's heatmap. Figure 5.19 shows the overall flow to obtain the covariance matrix from the CNN heatmaps.

### 5.6.2. PREDICTION

The first step of the filter is to generate sigma points $\boldsymbol{\chi}^{[i]}$ from the current state vector $\tilde{\boldsymbol{x}}_k$ by using the standard formulation of the UKF (Julier and Uhlmann, 2004):

$$\boldsymbol{\chi}_k^{[i]} = \begin{cases} \tilde{\boldsymbol{x}}_k & i = 0 \\ \tilde{\boldsymbol{x}}_k + \left(\sqrt{(N+\lambda)\boldsymbol{P}}\right)_i & i = 1,...,N \\ \tilde{\boldsymbol{x}}_k - \left(\sqrt{(N+\lambda)\boldsymbol{P}}\right)_i & i = N+1,\ldots,2N \end{cases} \tag{5.13}$$

where $N = 12$ is the dimensionality of the system and $\lambda$ is a scaling factor tuned offline. The sigma points are then propagated through the dynamic equations. For the translational motion, the ROE state is propagated assuming an unperturbed Keplerian orbit for the servicer spacecraft,

$$\boldsymbol{\delta\alpha}_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1.5n\Delta t & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \boldsymbol{\delta\alpha}_{k-1}, \tag{5.14}$$

where $n$ represents the mean motion of the orbit. For the attitude, each sigma sample $\boldsymbol{\delta p}^{[i]}$ is transformed to the error-quaternion $\boldsymbol{\delta q}^{[i]}$, which is used to generate the quaternion multiplicatively via $\boldsymbol{q}_k^{+,[i]} = \boldsymbol{\delta q}^{[i]} \otimes \boldsymbol{q}_k$, where $\otimes$ denotes the quaternion multiplication. The dynamics update of the quaternion representation of relative orientation assumes perturbation-free motion,

$$\boldsymbol{q}_k^- = \begin{bmatrix} \cos\left(\frac{1}{2}\left\|\boldsymbol{\omega}_{B/C}^C\right\|\Delta t\right) \\ \hat{\boldsymbol{\omega}}\sin\left(\frac{1}{2}\left\|\boldsymbol{\omega}_{B/C}^C\right\|\Delta t\right) \end{bmatrix} \boldsymbol{q}_{k-1}^+, \tag{5.15}$$

where $\Delta t$ is the time-step between filter calls and $\hat{\boldsymbol{\omega}}$ is the normalized angular velocity. Afterwards, $q_{k+1}^{-,[i]}$ is used to compute $dp_{k+1}^{-,[i]}$. The mean of $dp_{k+1}^-$ provided by the unscented transform is used to update the nominal state quaternion.

The relative angular velocity $\boldsymbol{\omega}_{B/C}^C$ is propagated via the following equation (Capuano et al., 2020):

$$\dot{\boldsymbol{\omega}}_{B/C}^C = \boldsymbol{R}_B^C(\boldsymbol{\tau}_B - \boldsymbol{\omega}_{B/E}^B \times \boldsymbol{J}_B\boldsymbol{\omega}_{B/E}^B) - \boldsymbol{J}_C^{-1}(\boldsymbol{\tau}_C - \boldsymbol{\omega}_{C/E}^C \times \boldsymbol{J}_C\boldsymbol{\omega}_{C/E}^C) - \boldsymbol{\omega}_{C/E}^C \times \boldsymbol{\omega}_{B/C}^C, \tag{5.16}$$

where $\boldsymbol{J}$ is the target's inertia matrix, $\boldsymbol{\tau}$ are the external forces, and $E$ denotes the Earth-Centered-Inertial (ECI) frame. Note that in the current implementation no control forces are assumed for neither the target nor the servicer, i.e. $\boldsymbol{\tau}_B = \boldsymbol{\tau}_C = \boldsymbol{0}$. After propagation, the sigma points are used to derive a mean state estimate and its associated error covariance,

$$\bar{\boldsymbol{x}}_k = \sum_{i=0}^{2N} w^{[i]}\boldsymbol{\chi}^{[i]}, \tag{5.17}$$

$$\boldsymbol{P}_k = \sum_{i=0}^{2N} w^{[i]}(\boldsymbol{\chi}^{[i]} - \bar{\boldsymbol{x}})(\boldsymbol{\chi}^{[i]} - \bar{\boldsymbol{x}})^T + \boldsymbol{Q}_k, \tag{5.18}$$

where the weights $w^{[i]}$ are a function of $\lambda, N$ and $\boldsymbol{Q}_k$ is the process noise covariance.

### 5.6.3. CORRECTION

The measurement update follows the projections described in Eqs. (2.1-2.2), in which the sigma points of the relative position $\boldsymbol{t}^C$ are derived from the ROE state given the knowledge of the servicer's orbital elements state and its attitude with respect to the ECI frame. Once the expected measurements $\boldsymbol{Z}^{[i]}$ given each sigma point are computed, the mean expected measurement and innovation covariance can be computed similar to the mean and covariance of the filter state,

$$\bar{\boldsymbol{z}}_k = \sum_{i=0}^{2N} w^{[i]} \boldsymbol{Z}^{[i]}, \tag{5.19}$$

$$\boldsymbol{S}_k = \sum_{i=0}^{2N} w^{[i]} (\boldsymbol{Z}^{[i]} - \bar{\boldsymbol{z}})(\boldsymbol{Z}^{[i]} - \bar{\boldsymbol{z}})^T + \boldsymbol{R}_k, \tag{5.20}$$

where $\boldsymbol{R}_k$ represents the measurement error covariance matrix. In the proposed system, $\boldsymbol{R}_k$ is a time-varying block diagonal matrix constructed with the heatmaps-derived covariances $\boldsymbol{C}_i$ in Equation (5.11),

$$\boldsymbol{R}_k = c_k \begin{bmatrix} \boldsymbol{C}_1 & & \\ & \ddots & \\ & & \boldsymbol{C}_n \end{bmatrix}, \tag{5.21}$$

where the scaling coefficient $c_k$ is described in Section 5.6.4. Notice that $\boldsymbol{C}_i$ can differ for each feature in a given frame as well as vary over time. Preliminary navigation results (Pasqualetto Cassinis et al., 2020; Pasqualetto Cassinis, Fonod, Gill, Ahrns, and Gil-Fernandez, 2021) already showed that such heatmaps-derived covariance matrix can capture the statistical distribution of the measured features and improve the measurement update step of the navigation filter.

At this stage, outliers are removed from the mean measurements by computing the Mahalanobis Distance $M_i$ between the $i$th feature and its corresponding filter innovation term $\boldsymbol{\Delta}_k^z$,

$$M_i = \sqrt{\boldsymbol{\Delta}_{k,i}^z \boldsymbol{S}_k^{-1} \boldsymbol{\Delta}_{k,i}^{z\,T}}, \tag{5.22}$$

where $\boldsymbol{\Delta}_k^z = \bar{\boldsymbol{z}} - \boldsymbol{h}(\bar{\boldsymbol{x}}_k)$ and $\boldsymbol{h}$ is the nonlinear transformation in Eqs. (2.1-2.2). The Mahalanobis distance is the distance between a point and a distribution (De Maesschalck et al., 2000). In this case, the point is the keypoint detected by the CNN in the image and the distribution is the reprojected feature and its associated covariance. The threshold $M_t$ to select outliers is determined by

$$M_t = \sqrt{-2\ln(p_m)}, \tag{5.23}$$

where $p_m = 90\%$ is the desired probability that a measurement would result in $M_i \geq M_t$, given that the correspondence between the detected and reprojected features is correct (Thrun et al., 2005). In other words, if $M_i \geq M_t$ it is highly unlikely that the $i$th keypoint correlates to its reprojection as predicted by the filter, and the feature can be rejected.

In this way, filter robustness can be improved during low visibility periods of the target spacecraft in which wrong CNN detections may occur. Notably, this feature rejection scheme reflects an important advantage of incorporating a navigation filter compared to relying solely on the CNN detection and PnP solver solution.

Finally, the corrected state estimate $\hat{\boldsymbol{x}}_k$ is obtained from the propagated state $\bar{\boldsymbol{x}}_k$, the innovation $\boldsymbol{\Delta}_k^z$, and the Kalman Gain $\boldsymbol{K}_k$,

$$\hat{\boldsymbol{x}}_k = \bar{\boldsymbol{x}}_k + \boldsymbol{K}_k \boldsymbol{\Delta}_k^z, \tag{5.24}$$

where $\boldsymbol{K}_k$ is a function of the state error and innovation covariance, and $\boldsymbol{K}_k \boldsymbol{\Delta}_k^z$ represents the state innovation.

After correction, the new attitude error $\boldsymbol{\delta p}$ is reset to zero at each iteration.

### 5.6.4. COVARIANCE ADAPTATION



As already reported in Chapter 3, the heatmaps-based covariance described in Section 5.6.1 already proved to accurately represent the measurements uncertainty of the CNN detections in synthetic images with ideal illumination conditions. However, a reliable representativeness of the heatmaps cannot always be guaranteed in realistic space images under challenging illumination conditions. Despite capturing the shape of the distribution, the magnitude of the heatmpas-derived covariance could indeed fail at representing the actual detection uncertainty of the CNN. This could lead the navigation filter to trust inaccurate features and ultimately diverge.

Referring to Figure 5.18, the above challenges are addressed by adaptively estimating the measurement error covariance through a new technique which leverages the heatmaps-based covariance together with existing covariance matching techniques. This approach is similar to the adaptive state noise compensation method for estimating the process noise covariance introduced in Stacey and D'Amico (2021).

By replacing the theoretical covariance of the innovations, also known as pre-fit residuals, in a Kalman filter with an empirical estimate, the measurement error covariance of the $i$th pair of pixel measurements at time step $k$ can be estimated as (K. A. Myers, 1974)

$$\hat{\mathbf{R}}_{i,k} = \boldsymbol{\Delta}_{i,k}^z \boldsymbol{\Delta}_{i,k}^{z^T} - \bar{\mathbf{S}}_{i,k}. \tag{5.25}$$

Here, the innovation $\boldsymbol{\Delta}_{i,k}^z$ is the difference between the true and expected $i$th pair of pixel measurement at time step $k$, taking into account all measurements through time step $k-1$. The matrix $\bar{\mathbf{S}}_k = \mathbf{S}_k - \mathbf{R}_k$ is computed from the sigma points passed through the nonlinear measurement models in Equation (5.20) (Thrun et al., 2005). The portion of $\bar{\mathbf{S}}_k$ corresponding to the $i$th pair of pixel measurements is $\bar{\mathbf{S}}_{i,k} \in \mathbb{R}^{2\times 2}$. Equation (5.25) assumes an optimal filter at steady state and that the pixel measurement errors are zero-mean. Typically, Equation (5.25) is averaged over some finite length sliding window of filter output. Such covariance matching techniques, also referred to as innovation based estimation, are not guaranteed to converge to the true estimate of the error covariance. However, they are widely used and have been shown to work well in practice (Fraser and

Ulrich, 2021; Karlgaard, 2010; Mohamed and Schwarz, 1999; K. Myers and Tapley, 1976; Sullivan and D'Amico, 2017).

Assuming that the pixel measurement errors are not correlated with each other and that the associated heatmaps are representative of the shape of their covariance, the measurement error covariance can be modelled as:

$$\mathbf{R}_{i,k} = c_k \mathbf{C}_{i,k}, \tag{5.26}$$

where $c_k$ is introduced to account for the uncertainty in the magnitude of the covariance. Using all the pixel measurements at a single filter call, a pseudo-coefficient $c_k^*$ can be estimated through a weighted least squares fit between the diagonal elements of the right hand side of Equation (5.26) and the corresponding elements of Equation (5.25). The resulting solution for $c_k^*$ is:

$$c_k^* = \arg\min_c ||\mathbf{W}_k^{-1/2}(\boldsymbol{A}_k c - \boldsymbol{b}_k)||^2 \tag{5.27}$$

$$= \frac{\boldsymbol{A}_k^T \mathbf{W}_k^{-1} \boldsymbol{b}_k}{\boldsymbol{A}_k^T \mathbf{W}_k^{-1} \boldsymbol{A}_k}. \tag{5.28}$$

The vector $\boldsymbol{b}_k$ is the concatenation of all the main diagonal elements of the covariance matching estimates $\hat{\mathbf{R}}_{i,k} \in \mathbb{R}^{2\times 2}$ for all $i$ at time step $k$. The vector $\mathbf{A}_k$ contains the corresponding diagonal elements of each $\mathbf{C}_{i,k}$.

The weighting matrix $\mathbf{W}_k$ is chosen as the theoretical covariance of $\boldsymbol{b}_k$ such that covariance matching estimates with less uncertainty have a greater influence on the solution of $c_k^*$. The covariance between any two elements of $\boldsymbol{b}_k$ is

$$\mathrm{Cov}(\hat{R}_{i,k}^{\alpha}, \hat{R}_{j,k}^{\beta}) = \mathrm{Cov}(\Delta_{i,k}^{\alpha^2}, \Delta_{j,k}^{\beta^2}) \tag{5.29}$$

$$= \mathrm{E}[\Delta_{i,k}^{\alpha^2}\Delta_{j,k}^{\beta^2}] - \mathrm{E}[\Delta_{i,k}^{\alpha^2}]\mathrm{E}[\Delta_{j,k}^{\beta^2}] \tag{5.30}$$

$$= 2\mathrm{E}[\Delta_{i,k}^{\alpha}\Delta_{j,k}^{\beta}]^2. \tag{5.31}$$

The scalar $\hat{R}_{i,k}^{\alpha}$ is either the first or second element of the main diagonal of $\hat{\mathbf{R}}_{i,k}$ as denoted by $\alpha$. Equation (5.31) is just an element of $\mathbf{S}_k$ squared multiplied by two. Thus,

$$\mathbf{W}_k = \mathbf{S}_k^{\circ 2}, \tag{5.32}$$

where the factor 2 from Equation (5.31) has been dropped because it does not change the result of Equation (5.28). The Hadamard power, $^\circ$, denotes an element-wise power. Equation (5.32) assumes that $\boldsymbol{b}_k$ is ordered in the same way as the pixel measurements are ordered in the measurement vector provided to the filter. For computational efficiency, the weighting matrix is approximated as diagonal by setting all off-diagonal elements to zero. As a result, the inverse of the weighting matrix required in Equation (5.28) can be computed element-wise along the main diagonal.

Assuming an optimal filter, the variance of the weighted least squares estimate (Gill and Montenbruck, 2012) is:

$$P_{c_k^*} = \text{Var}(c_k^*) \tag{5.33}$$

$$= (\boldsymbol{A}_k^T \mathbf{W}_k^{-1} \boldsymbol{A}_k)^{-1}. \tag{5.34}$$

The final estimate of $c_k$ is obtained by combining the $c^*$ estimates over a sliding window of length N through

$$c_k = \arg\min_c ||\bar{\mathbf{W}}_k^{-1/2}(\bar{\boldsymbol{A}}_k c - \bar{\boldsymbol{b}}_k)||^2 \tag{5.35}$$

$$= \frac{\bar{\boldsymbol{A}}_k^T \bar{\mathbf{W}}_k^{-1} \bar{\boldsymbol{b}}_k}{\bar{\boldsymbol{A}}_k^T \bar{\mathbf{W}}_k^{-1} \bar{\boldsymbol{A}}_k}, \tag{5.36}$$

where

$$\bar{\boldsymbol{A}}_k = \mathbf{1}_{N\times 1}, \quad \bar{\boldsymbol{b}}_k = [c_k^* \ \dots \ c_{k-N+1}^*]^T \tag{5.37}$$

$$\bar{\mathbf{W}}_k = \begin{bmatrix} P_{c_k^*} & 0 & \dots & 0 \\ 0 & P_{c_{k-1}^*} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & P_{c_{k-N+1}^*} \end{bmatrix}. \tag{5.38}$$

Upper and lower inequality constraints are easily added to Equation (5.35) by setting the estimated $c_k$ equal to any constraint that is violated. Note that some lower bound greater than zero should always be included to guarantee that the measurement error covariance is positive definite. Any filter calls where no pixel measurements are provided to the filter are excluded from the sliding window. The weighting matrix in Equation (5.38) is assumed diagonal because the innovations are not correlated in time for an optimal filter at steady state (Kailath, 1968; Mehra, 1972). After each measurement update, $c_k^*$, $P_{c_k^*}$, and $c_k$ are computed through Eqs. (5.28), (5.34), and (5.36) respectively. The resulting $c_k$ is used in the following measurement update to compute the measurement error covariance of each pixel measurement through Equation (5.26).

## 5.7. SIMULATIONS

The investigated rendezvous scenarios involve the Mango and Tango spacecraft from the PRISMA mission (D'Amico et al., 2013). The simulations take place in Low Earth Orbit (LEO) with initial ROEs listed in Table 5.4. ROE 1 is purely an along-track separation describing a standard V-bar hold point, whereas ROE 2 describes a mid-range hold point (Sharma and D'Amico, 2017). The reference truth motion of each spacecraft is numerically propagated with 1 second interval using rigorous force models including GRACE's GGM05S geopotential of order and degree 120 (Ries et al., 2016), NRLMSISE-00 atmospheric model (Picone et al., 2002), analytic lunisolar third-body gravity (Gill and Montenbruck, 2012), and solar radiation pressure (SRP) including a conical Earth shadow model. Moreover, the spacecraft attitude is perturbed via analytical gravity-gradient

Table 5.4: Initial mean servicer orbital elements and reference relative trajectories parametrized in ROE space (Sharma and D'Amico, 2017).

| Servicer Orbit | $a = 7078.1$ [km] | $e = 0.001$ | $i = 98.2°$ | $\Omega = 189.9°$ | $\omega = 0°$ | $M_0 = 0°$ |
|---|---|---|---|---|---|---|
| Initial ROE | $a\delta a$ [m] | $a\delta\lambda$ [m] | $a\delta e_x$ [m] | $a\delta e_y$ [m] | $a\delta i_x$ [m] | $a\delta i_y$ [m] |
| ROE 1 | 0 | -8 | 0 | 0 | 0 | 0 |
| ROE 2 | -0.25 | -8 | 0 | 0.15 | 0 | -0.15 |

(a) ROE 1

(b) ROE 2

Figure 5.20: First five HIL images for ROE 1 (a) and ROE 2 (b) trajectories. The images were generated every 30 s to represent limited-on board processing power.

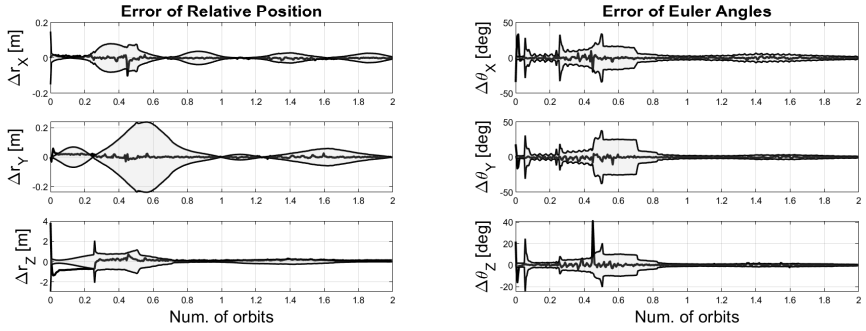
(Wertz, 1978), air drag, and SRP effects. The satellite parameters for these forces match the modeled parameters of the Mango and Tango spacecraft specified in D'Amico (2010). The servicer's angular velocity is initialized to always align the camera boresight with the along-track direction, whereas the target's initial angular velocity about its principal axes is set to $\boldsymbol{\omega}_0 = [1 \quad 0 \quad 0]^T (°/s)$ for ROE 1 and $\boldsymbol{\omega}_0 = [0 \quad 0.4 \quad -0.6]^T (°/s)$ for ROE 2. The images are captured every 5 seconds. However, a measurement acquisition time of 30 seconds is used in this work, in order to represent a severely limited on-board processing power. The synthetic images are created using the same OpenGL-based graphics renderer used to create SPEED+, whereas the realistic images are generated in TRON following the procedure described in Section 5.3. Figure 5.20 shows a montage of the HIL images for both ROE trajectories.

## 5.8. RESULTS

In this section, the navigation results are presented for the ROE 1 and ROE 2 trajectories. Similar to the pose estimation error metric introduced in Section 5.5, the norms of the estimated translational and rotational velocities are compared to their ground truth values to return the estimation error. Table 5.5 reports the obtained navigation results for both synthetic and HIL scenarios in terms of mean error at steady state. The mean values after two orbits are computed across a time interval of 600 s at steady state.

Table 5.5: Mean navigation results after two orbits for ROE 1 and ROE 2 trajectories, computed across a time interval of 600 s at steady state. Results are reported for both synthetic and HIL scenarios.

| Trajectory | Measurements | $E_T$ [m] | $E_R$ [m] | $E_v$ [m/s] | $E_\omega$ [°/s] |
|---|---|---|---|---|---|
| ROE 1 | `synthetic` | 0.13 ± 0.02 | 0.73 ± 0.37 | 1.2E-11 ± 2E-12 | 0.008 ± 0.005 |
| | `lightbox` | 0.25 ± 0.06 | 14 ± 11 | 1.3E-11 ± 1.6E-12 | 0.5 ± 0.4 |
| ROE 2 | `synthetic` | 0.02 ± 0.02 | 0.97 ± 0.85 | 8.5E-12 ± 1E-12 | 0.01 ± 0.005 |
| | `lightbox` | 0.38 ± 0.02 | 9 ± 7.5 | 4.9E-11 ± 6.8E-12 | 0.3 ± 0.2 |




(a) (b)

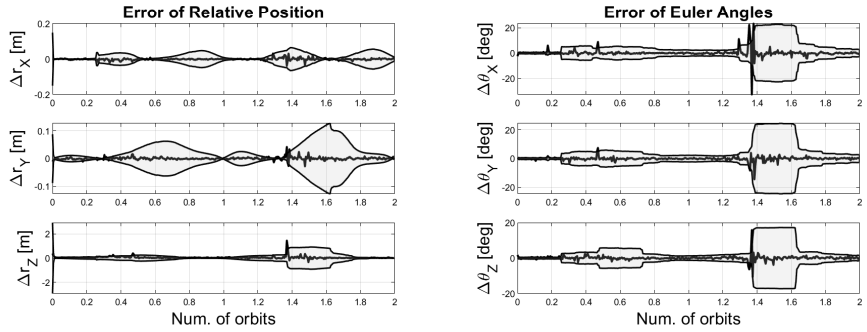
Figure 5.21: Pose Estimation Error in the ROE 1 synthetic scenario. The ROE state is converted to position (left) and the quaternion state is converted to 3-2-1 Euler angles (right). The errors are shown along with the $3\sigma$ error bound derived from the corresponding diagonal entries of the filter error covariance matrix.

### 5.8.1. SYNTHETIC SCENARIOS

Figures 5.21-5.22 show the navigation performance on both relative trajectories when synthetic images are used as measurements. Results are reported for the pose estimation error along with the standard deviation derived from the state covariance matrix. In these scenarios, centimeter-level position error and degree-level attitude error are achieved at steady state after two relative orbits. Also, notice that the state covariance increases at approximately half of each relative orbit before quickly decreasing. In these periods, some of the CNN detections become inaccurate due to the challenging near-eclipse conditions in the target's visibility. However, thanks to the adaptive heatmaps-based measurement error covariance the filter can capture the detection uncertainty and reflect it in a larger state covariance. This is a desirable behaviour as it prevents the filter from diverging due to an inaccurate representation of the measurements uncertainty.

### 5.8.2. TRON SCENARIOS

As mentioned in Section 5.6.4, the magnitude of the heatmaps-based covariance is not guaranteed to reflect the actual measurements uncertainty in realistic imagery, especially



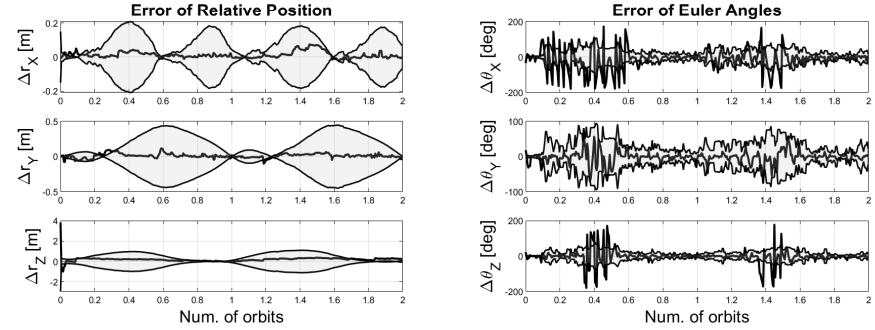

(a) (b)

Figure 5.22: Pose Estimation Error in the ROE 2 synthetic scenario. The ROE state is converted to position (left) and quaternion state to 3-2-1 Euler angles (right). The errors are shown along with the $3\sigma$ error derived from the corresponding diagonal entries of the state covariance matrix.



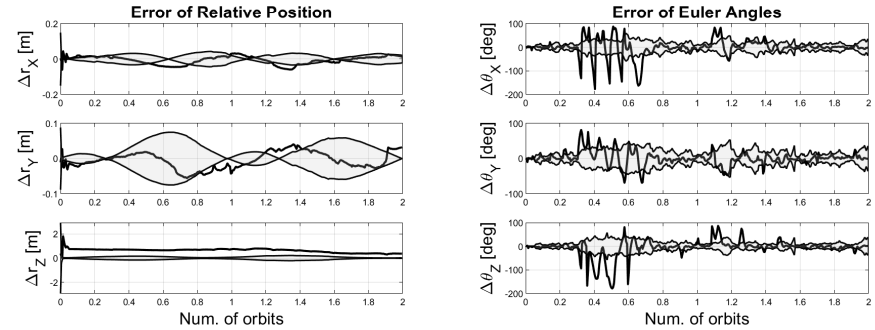
under highly challenging illumination conditions. As such, the state covariance could be updated with an inaccurate measurement error covariance (Equation (5.20)), leading to the divergence of the filter immediately after initialization. This was observed for both ROE trajectories when a simple heatmaps-based covariance was used. However, the achieved performance of the filter can benefit from the proposed covariance adaptation scheme. After the estimated scaling factor is introduced (Equation (5.26)), the filter shows an increase in robustness towards inaccurate measurements and does not diverge. Figures 5.23-5.24 show the navigation performance on both relative trajectories when the HIL images are used. Thanks to the covariance adaptation, centimeter-level position error is achieved after two relative orbits, which compares well with the results in the synthetic scenario. However, a larger uncertainty in the attitude estimate can be seen in both trajectories, with steady-state mean errors $E_T \approx 10^\circ - 15^\circ$. A degraded performance in the relative attitude due to challenging HIL images was already observed in Section 4.6.3, suggesting that the domain shift affects the estimate of the rotational state more than the translational one. This property was deemed to be related to the fact that for large domain gaps the network can confuse similar corners and lead to a wrong 2D/3D point correspondence. As such, the P$n$P solver can wrongly estimates the object attitude despite a correct estimate of the position. Nevertheless, the velocity estimates listed in Table 5.5 indicate that the proposed system can estimate the relative rotational velocity with relatively small errors.

Overall, the gained filter robustness highlights the benefits of adaptively correcting the measurement error covariance. Notably, the proposed covariance adaptation scheme was applied to a more standard, gradient-based covariance method (Cui et al., 2019) which does not exploit the CNN heatmaps, in order to compare its performance with the heatmaps-based method. Preliminary results showed a much worse navigation performance with mean attitude errors $E_R > 100^\circ$ after one relative orbit, suggesting that the proposed heatmaps-based representation correlates with the CNN uncertainties




(a) (b)

Figure 5.23: Pose Estimation Error in the ROE 1 TRON scenario.




(a) (b)

Figure 5.24: Pose Estimation Error in the ROE 2 TRON scenario.

better than a gradient-based method. This indicates that not only the adaptive covariance is essential for filter robustness, but also that a proper representation of the measurements uncertainty is required beforehand.

## 5.9. CHAPTER CONCLUSIONS

This Chapter introduces the on-ground validation of an adaptive Convolutional Neural Network (CNN)-based navigation filter for the pose estimation of an uncooperative spacecraft by an active servicer spacecraft equipped with a monocular camera. First, the performance of the proposed system is evaluated at a pose estimation level by testing the adopted CNN on the realistic space imagery of the SPEED+ dataset, captured from Stanford's robotic Testbed for Rendezvous and Optical Navigation (TRON). By adopting a novel data augmentation pipeline centered on light augmentation, the system is proven capable of bridging the domain gap between the synthetic training images and the HIL test images, showcasing the benefits of generalizing the illumination conditions during

training. Specifically, pose errors $E_T < 0.1 \left\| \boldsymbol{t}^C \right\|$, $E_R < 10°$ and $E_T < 0.1 \left\| \boldsymbol{t}^C \right\|$, $E_R < 13°$ are achieved in 80% of the `lightbox` and `sunlamp` subsets, respectively. Furthermore, by assessing the pose estimation performance on both TRON images and actual space imagery from the PRISMA mission, the data augmentation method is shown to be robust against changes in the target domain. Besides, the implementability of the proposed HRNet on a Myriad X processor is showcased by achieving a comparable performance to the GPU implementation with a x6 better inference per Watt.

At a navigation filter level, the system is validated by assessing the performance of the proposed CNN-based Unscented Kalman Filter (UKF) on both synthetic and HIL images of two representative rendezvous trajectories around the target spacecraft. Results on the synthetic scenarios indicate that the system can accurately estimate the relative state with centimeter- and sub-degree-level pose errors, thanks to the heatmaps-based representation of the CNN detection uncertainty. In the HIL scenarios, the inclusion of an adaptation scheme for the measurement error covariance returns centimeter-level position errors and moderate attitude accuracies ($E_R \approx 10° - 15°$) at steady-state, preventing the filter from diverging during periods of low measurement accuracy. Remarkably, these results suggest that a proper representation of the measurements uncertainty combined with an adaptive measurement error covariance is key in improving the navigation robustness.

# 6

# CONCLUSION

## 6.1. MAIN FINDINGS AND CONCLUSIONS

The estimation of the relative pose of an uncooperative target with respect to a servicer spacecraft is a critical navigation task during proximity operations such as fly-around, inspections and close-approach in Active Debris Removal (ADR) and On-Orbit Servicing (OOS) missions. This thesis analyzed the challenges of a monocular vision-based relative pose estimation system, in which a single monocular camera is used as a navigation sensor due to its reduced mass, power consumption and system complexity compared to systems based on active sensors or stereo cameras. Building on the characteristics and limitations of existing methods, this thesis investigated the feasibility of a Convolutional Neural Network (CNN)-based relative pose estimation method and its system interfaces with a navigation filter, with special focus given to the on-ground validation of the proposed navigation pipeline with realistic laboratory images of target spacecraft. This section describes the main findings of this thesis associated to the research questions introduced in Chapter 1.

**Research Question #1**

What are the characteristics and limitations of monocular vision-based systems for the pose estimation of uncooperative spacecraft?

The key to identify the characteristics and limitations of a relative pose estimation system lies in dissecting such system into its main tasks. From a high level perspective, the major limitations of a monocular vision-based system are traced back to the image processing, which is in charge of extracting useful information from a 2D image. Although standard feature detectors are proven to perform with high accuracy and robustness in ideal scenarios, their performance can drastically decrease in adverse scenarios, such as near-eclipse conditions, unfavourable target reflections and Earth in the background. The detection of variable edges and corners on the target object due to changing view

conditions can result in a lengthy 2D/3D correspondence and affect the pose estimation accuracy and robustness. Moreover, the extraction and matching of features across subsequent images can be jeopardized by large inter-frame rotations of the target as well as large variations in illumination conditions. Altogether, these aspects highlight a lack of robustness of standard image processing algorithms against adverse viewing conditions of the target spacecraft. In this context, the introduction of CNNs in the framework of relative pose estimation, and their use as keypoint detectors prior to the actual pose solver, led to a considerable improvement in the feature detection performance. One of the main characteristics of CNN-based feature detectors lies in the pre-selection of a fixed set of keypoint features during training, which avoids a lengthy 2D/3D correspondence. Furthermore, a proper training results in high robustness against the same adverse viewings of the target that resulted in a poor performance for the more standard, CNN-free methods. Then, at a navigation filter level it was discovered that filter performance is, from a high-level perspective, tied to filter robustness and filter optimality. In particular, filter optimality can be linked to the accuracy of the selected pose estimation method prior to the actual filtering. Conversely, filter robustness is highly affected by the challenges in representing the measurements uncertainty of monocular sensors, due to highly varying orbital conditions. Both optimality and robustness are key performance indicators when in close-proximity with an uncooperative spacecraft, and they have to be properly traded-off when selecting a pose estimation system for ADR/OOS missions.



Beyond the characteristics and limitations of each subsystem, the most important aspect that was identified in a monocular-based relative pose estimation system is represented by a dire need to establish a validation framework on-ground with representative images as a surrogate of actual space imagery. Although such validation is of high importance regardless of the selected pose estimation method, this need is critical if a CNN-based method is selected, due to the so-called domain shift problem. Specifically, synthetically-trained CNNs tend to decrease their performance on realistic imagery, if the adopted training dataset fails at representing realistic illumination conditions, image noise and other artifacts present in the real environment. To cope with these limitations, data augmentation or domain adaptation methods need to be assessed and extensively tested on large datasets.

**Research Question #2**

Which pose estimation methods return the robustness and accuracy required for the relative pose estimation of an uncooperative spacecraft?

The better performance of keypoints-based CNNs over end-to-end CNNs observed during the Spacecraft Pose Estimation Dataset (SPEED) competition already suggested that a CNN-based keypoint detector in combination with a standard P$n$P solver represents a promising architecture for a robust and accurate monocular-based relative pose estimation. This conclusion also stems from the challenges identified in the vast majority of CNN-free pose estimation methods, in which the overall pose estimation accuracy tends to be jeopardized by the complications involved in the detection of features on the target spacecraft in adverse illumination conditions. To get further insights on the performance of a keypoint based, CNN-based relative pose estimation method, an analysis

was carried out on synthetic images of the Envisat spacecraft, representing typical close-proximity approach phases of an ADR mission. At a detection level, a light, Single-stack hourglass network was compared with the heavier and more accurate HRNet, in order to investigate the possibility to have smaller CNNs and decrease the computational effort during image processing. In both networks, it was discovered that thanks to the capability of CNNs to extrapolate keypoints location during training, occluded and partially-visible features could be detected. In this way, the pose estimation accuracy could be increased by providing a reliable set of features to the pose estimation solver, even in near-eclipse illumination conditions.

Building on the gained insights at a CNN detection level, it was further observed that the higher resolution and number of parameters of the HRNet provide a higher pose estimation accuracy and robustness when compared to a Single-stack hourglass. Notably, the performance of light networks such as the Single-stack hourglass is expected to drastically decrease when tested on more realistic images of target spacecraft, mostly due to the challenge for its smaller number of convolutional layers to handle previously unseen textures and illumination conditions. This delineates a key aspect at image processing level, in which an increase in robustness towards highly adverse illumination scenarios must be traded-off with an increase in number of parameters of a CNN and, in turn, with an increase in computational effort for feature detection. This is crucial when the proposed pose estimation methods have to be implemented in space-representative processors. In this context, a novel method was proposed to capture the detection uncertainty from the CNN heatmaps for each feature. This was done in order to improve the estimation robustness by providing a reliable statistical interpretation of feature uncertainty to the P$n$P solver, while maintaining a small number of CNN parameters and facilitate the implementation of light CNNs on space-representative processors. Although three selected scenarios suggested that a heatmaps-based representation of feature uncertainty can aid pose estimation when the heatmaps effectively represents the CNN uncertainty, results on the entire test dataset indicated that less reliable representations of feature uncertainty are likely to occur in adverse illumination scenarios. This indicates that a heatmaps-based feature detection uncertainty could be hard to be exploited at a pose estimation level.



**Research Question #3**

Which characteristics are critical in a navigation filter for the relative pose estimation of an uncooperative spacecraft?

The use of a navigation filter for the estimation of the relative state of an uncooperative spacecraft with respect to a servicing spacecraft is essential in close-proximity operations which characterize ADR and OOS missions. A navigation filter provides the estimation of translational and rotational velocities that add to the actual pose estimate from direct pose estimation. Besides, it returns estimates at higher frequencies and can fuse the monocular camera measurements with its propagation of the system dynamics. One of the main characteristics of a monocular vision-based relative navigation filter lies in the way the measurements are represented in the filter, either by directly feeding detected features (tightly-coupled) or by processing the features into so-called pseudomeasurements

of the relative pose with a pose solver (loosely-coupled). When a CNN is used at a feature detection level, a tightly-coupled architecture seems the most reasonable solution as it is more practical to derive the measurement error covariance for detected features rather than from the estimated pose. In this context, filter robustness is proven to be highly affected by the representability of the measurement error covariance associated to the CNN measurements. This thesis introduced a novel method in which the measurement error covariance is computed directly from the CNN heatmaps. Results with synthetic images show that a CNN-based navigation filter can accurately estimate the relative state, when the measurement uncertainty is derived from heatmaps. Moreover, it was discovered that the characteristic of having a tightly-coupled navigation filter can be an asset in coupling together the estimation of the translational and rotational states. Although in a generic sense a tightly-coupled approach could be less desirable than a loosely-coupled one when the target is tumbling and uncooperative, due to the highly variable number of features to track, the use of a CNN as a detector of a predefined set of features can provide a much more stable number of features. In summary, a system validation of the proposed relative navigation system was carried out on simulated synthetic trajectories. Results indicate that the proposed system, made of a CNN at a feature detection step, a heatmaps-based computation of the measurement error covariance, a pose estimation solver to compute the initial state of the filter, and a tightly-coupled filter architecture to estimate the full relative state, represents a viable solution which can return the accuracy and robustness required in ADR and OOS missions.



**Research Question #4**

How can the performance of monocular vision-based relative pose estimation systems on actual space imagery be improved?

The on-ground validation of a monocular vision-based relative pose estimation system on images representative of actual space imagery is paramount to assess the expected performance of said system in orbit. Although a synthetic validation of the system architecture as a whole can return valuable insights on the applicability of the system on the intended ADR and OOS missions, the performance at an image processing level can undergo a considerable drop from a synthetic rendering domain to the actual space domain, jeopardizing the understanding on the expected performance in orbit. This challenge is observed for any visual-based relative pose estimation system, and it is referred to as the domain shift problem in CNN-based systems. The domain shift problem states that a synthetically-trained CNN has a considerable performance drop when tested on images with previously unseen textures, light conditions, and backgrounds. To propose a solution to this problem and to extend the synthetic validation of the proposed CNN-based relative pose estimation system, this thesis investigated the system performance on Hardware-In-the-Loop (HIL) laboratory images. From a CNN training perspective, the inclusion of textures and light randomization was found to considerably improve the feature detection accuracy and robustness under a large variety of viewing conditions, such as adverse reflections of the target, occlusion of parts of the target body, presence of Earth in the background, and near-eclipse conditions. In turn, this led to an increase of accuracy and robustness at a pose estimation level. Notably, the validity of the HIL images

as a surrogate of actual space imagery was assessed by comparing the pose estimation results on both HIL images and actual flight data from the PRISMA mission. Furthermore, the performance analysis of the adopted CNN model on the HIL images of the SPEED+ dataset was extended by comparing the pose estimation accuracy and inference time on a GPU and on the low-power Myriad X processor. Despite on a limited set of images with highly adverse illumination conditions the performance on Myriad X can considerably decrease, results on the entire datasets indicate a comparable pose estimation accuracy with only a x2 slower inference on Myriad X.

At a navigation filter level, filter robustness was improved by introducing both a feature rejection scheme based on the Mahalanobis distance and a novel filter innovation-based adaptive scheme to scale the measurement error covariance. This latter approach combines the proposed heatmaps-based representation of the measurement error covariance with a covariance-matching representation based on filter innovation. This was proven to prevent the filter from diverging during periods of low measurement accuracy, suggesting that a proper representation of the measurements uncertainty combined with an adaptive measurement error covariance is key in improving the navigation robustness.

## 6.2. KEY INNOVATIONS AND CONTRIBUTIONS OF THESIS

The main objective of research work presented in this thesis was to develop and validate a robust and accurate monocular camera-based relative pose estimation system compliant with navigation requirements of ADR/OOS missions. Throughout this thesis, the proposed CNN-based relative pose estimation system was investigated as a whole as well as in relation to the main critical tasks of each independent subsystem, with special focus on the interface between the image processing algorithm, the pose estimator, and the navigation filter. This strategy resulted in four main novel contributions:



1. The introduction of a **heatmaps-based representation of feature detection uncertainty** applicable to CNN-based relative pose estimation systems, presented in Chapter 3.

2. The creation of a **calibration framework for ESTEC's GRALS testbed** to validate the performance of visual-based relative pose estimation systems on representative lab imagery, presented in Chapter 4.

3. The introduction of a **training data augmentation pipeline centered on texture and light randomization** to bridge the domain gap in CNN-based systems, presented in Chapter 4 and Chapter 5.

4. The introduction of a **filter innovation-based adaptive scheme to scale the measurement error covariance** and improve the statistical knowledge of the measurements within the navigation filter, presented in Chapter 5.

These developments were primarily centered around the concept that the validation of a monocular-vision based relative pose estimation system shall focus on the system performance on realistic space imagery. In this context, the identified challenges and the proposed solutions were presented in relation to the domain gap of CNN-based relative pose estimation systems.

## 6.3. RECOMMENDATIONS FOR FUTURE RESEARCH

In the framework of monocular vision-based relative pose estimation, several challenges stem from the complex tasks of each subsystem, ranging from feature detection at image processing level to the filter estimate of the full relative state. The main directions for future research in the field identified in this thesis are listed as follows.

**Towards highly performing network architectures and domain adaptation** The use of CNNs as keypoint detectors in monocular vision-based relative pose estimation systems has led to a breakthrough into the pose estimation accuracies that can be achieved under severely adverse views of the target object. Although effective offline training schemes such as the data augmentations proposed in this thesis are expected to improve their accuracy and robustness, a pletora of other options exists to improve their detection performance and extend their range of applicability. Certainly, one of the main aspects driving the CNN performance on actual space imagery is represented by the resolution of their network architectures. CNNs architectures have moved from simpler and lighter architectures such as the AlexNet and the hourglass network, to the most recent HRNet, with the latter improving on the network resolution by adopting multi-scale fusions across parallel convolutions. In this context, some recent studies proved that feature detection on unseen image domains can be greatly improved by having a shared multi-scale feature encoder and multiple prediction heads performing different related tasks, such as keypoints prediction, direct pose regression, and binary segmentation (Park and D'Amico, 2022b). Further studies in this direction are encouraged to extend the performance of CNN-based systems at an image processing level. Furthermore, another aspect that stands out as an alternative to training data augmentation is represented by domain adaptation. As already mentioned in this thesis, domain adaptation is an effective tool to improve the network performance by adapting the CNN on a specific target domain post training. Recent efforts in this direction demonstrated that by refining the parameters of the network layers via an online domain refinement in a self-supervised way, the CNN performance on the target domain images can be improved without their pose labels and with minimal computational cost (Park and D'Amico, 2022b).

**Perfecting the interface between CNNs and navigation filters** A critical aspect investigated in the proposed relative pose estimation system resides in the interface between the CNN-based relative pose estimation method and the navigation filter. Although this thesis focused on the modeling of the error covariance associated to the CNN measurements and proposed both an adaptive and a rejection scheme to cope with faulty measurements, it has not been fully addressed yet how the system dynamics can be exploited to aid the CNN detection on a sequence of images. More specifically, it is still unclear how the temporal dynamic behavior of CNNs can be improved when reliable feature tracking is demanded in low visibility of the target object. Future research should focus on the implementation of recurrent neural networks as well as on the inclusion of a feedback loop that allows the filter solution at a previous step to be used by the CNN during feature detection. This latter aspect should improve the capability of the CNN to track features thanks to the added knowledge on their inter-frame motion returned by the previous filter estimate.

**A new frontier in the on-ground validation** The methodology applied in this thesis to validate the proposed monocular vision-based relative pose estimation system follows the current trend of simulating representative HIL images in a laboratory environment as a surrogate of flight images. However, the use of such robotic testbeds typically results in high maintenance costs, workforce and required space which are not feasible for many research institutions. To foster a more flexible, reliable and versatile generation of realistic space imagery, Generative Adversarial Networks (GAN) (Creswell et al., 2018) should be investigated as an alternative to extensive laboratory calibrations and image acquisition campaigns. In this concept, realistic laboratory or actual space images are assumed to be available, either from an already existing testbed or from an already flown mission. Together with pre-generated synthetic images of the desired target object, these realistic images are used by the GAN to simulate realistic effects into the synthetic renderings and bridge the domain gap between synthetic rendering and actual space imagery. In this way, the simulation of HIL images can be avoided without jeopardizing the need to have realistic imagery during the on-ground validation. Surely, extensive research is needed in this context to guarantee an adequate actual space representativeness of the GAN-generated images.

**Relative pose estimation of unknown targets** This thesis investigated several challenges involved in the model-based relative pose estimation of an uncooperative known target, for which a wireframe model is available offline prior to the actual estimation and the target parameters mass and inertia are assumed to be known. As a result of these assumptions, the proposed system does not estimate unknown target properties, nor it recreates a representative model of the target online. However, the planned ADR and OOS missions are expected to deal with uncooperative unknown targets (i) whose geometry has most certainly deviated from the wireframe representation available prior to their launch, and (ii) whose mass and inertia properties can considerably differ from their initial values as a result of collision with space debris or due to failures that occurred during their orbit lifetime. Moreover, an unknown target can pose additional challenges to the CNN at an image processing level, due to a larger domain gap between its synthetic rendering and the actual space images acquired during the mission. To cope with these aspects, the results of this thesis shall be extended towards a navigation pipeline which can reconstruct a more representative wireframe model of the target in orbit, while including the unknown mass and inertia properties of the target in the full state vector of the navigation filter. From a mission design perspective, these tasks can be accomplished during a so-called *inspection phase*, in which the servicer spacecraft is put in a passively safe trajectory around the target. During this phase, the knowledge of the target state can be enhanced by refining the estimate of the target geometry and generate unlabeled images that can be used to perform domain adaptation and improve the CNN performance at later phases of the mission.

# ACKNOWLEDGEMENTS

*'Caminante, no hay camino, se hace camino al andar (wayfarer, there is no path, you make the path as you go)'*. This book starts with this quote from Antonio Machado, a quote that every pilgrim has heard multiple times on his/her way to Santiago de Compostela. When I started the Camino de Santiago back in 2018, I did not expect any religious or spiritual experience. To my eyes, the Camino represented an incredible opportunity to do what I, like the majority of the Venetians I know, like doing more than anything else: talking and walking. To my surprise, I very soon realized that the journey was instead much more than a mainstream experience. I came to discover that the most intense and true beauty lies in simple things; that receiving a hug is the only thing you need in order to call any day a beautiful day. Most importantly, I opened up to the realization that strong emotions are what keep us alive, whether they are good or bad. But probably the moment that still resonates with me the most to these days is when the owner of a tiny hostel in the middle of the Asturian mountains wrote in my travel diary that the true *camino* (path) starts, and not ends, in Santiago.

I am starting the acknowledgments with this experience because the camino is, in many ways, what has shaped my current self the most. Ironically, I believe that the path I undertook during those two weeks in Spain prepared me for the much longer Ph.D. path, perhaps by reminding me that the process of learning could be (and most of the time is) as irregular and slopy as the mountains in Northern Spain.

First and foremost, I would like to thank my promotors and daily supervisors at TU Delft: Prof. Eberhard Gill, Robert Fonod and Alessandra Menicucci. To them I express my deepest gratitude for the fruitful discussions, the vital supervision throughout the entire thesis, and their kindness and availability. Also, special thanks goes to the Ph.D. community I was part of during these long years (Stefano, Fiona, Marsil, Mario, Victor, Erdem, Joshua, Mario, Juan, Gitte and Dennis, Livio, Rashika, Ruipeng and Dora): Sharing my passion for whisky and hearing about your amazing researches while having some genuine laughs was such an important part of my days in Delft.

My Ph.D. experience has been a joint collaboration of several institutions, Agencies and industries. Without any of these partners, the final outcome of the thesis would have suffered greatly. For this reason, I would like to first of all thank all my direct supervisors and colleagues at the European Space Agency (Jesus, Manolo, Joris, Irene, Antoine, Elio, Olivier, Massimo, and the entire GNC team), with special thanks to Martin: our endless times alone in the darkness of GRALS, the deep philosophical discussions about life, the epic climbing trips to Belgium and your never-ending kindness and openness are probably what made me both a better researcher and a happier person. I will keep the moments we shared together with me forever. Needless to say, my entire Chapter 4 would not exist without your unbelievable patience and skills.

I would also like to thank my supervisor Ingo Ahrns together with all my colleagues from Airbus Defence and Space Bremen (Christoph, Verena, Nick, Ralf and Torsten),

for their supervision and guidance but also for the fun times at the Beer gardens. My experience in Germany was so short and frenetic, yet you all managed to make it so exciting and fruitful.

A big part of my experimental validation was carried out at Stanford Rendezvous Laboratory. This would have never been possible without the guidance and support of Prof. Simone D'Amico. His vision behind the establishment of the lab and his priceless supervision and knowledge are a fundamental component of my 6 months stay at Stanford University. Likewise, the help and support I received from his Ph.D. students was so crucial that it can hardly be quantified. A special thanks goes to Jeff for operating the TRON testbed while sharing his research ideas and outputs, and to Nathan for sharing his incredible knowledge on estimation theory and covariance adaptation, but also to the entire SLAB group (Tommaso, Justin, Kaitlin, Matthew and Shane) for making my days in the basement so valuable and memorable. Incidentally, my 6 months in California have been one of the greatest experience I had in my entire life, and this would have never been the case was it not for the amazing people I met on the way and shared experiences with. To the Rainbow Mansion community (Hemil, Roberto, Witold, Damiana, Vanessa, Nick, Dan(s), Hannah and Jenica): you truly are incredible. I want you all to know that you made my research period in the Bay Area so gold by enriching my pre/after work hours. To Ted, Melanie, Erin, Kris, Jake, Daniel, Camille and Bruce: travelling to the most remote parts of California with just a rope, a harness and a tent has been priceless. You truly are the confirmation that the reason why I love climbing lies in the climbers I meet. It is not that hard to believe that I have found eye-opening research inspirations during our endless laughs, the screams while projecting a hard route, the endless driving with Led Zeppelin and Pink Floyd, or during our awakenings on a crashpad in Yosemite.

A big part of my Ph.D. was affected by the recent pandemic and by the resulting social distancing. I would have never been able to survive this period without the constant presence of my beloved Leiden buddies. To Bastien, Martin, David, Jose, Laura, Joris, Elena and Sara: we nailed it! Our impulsive trips to Belgium and the Canary islands were for me like a second rebirth. I believe that in our obsessiveness for climbing walls (and houses), dancing Sirtaki and sharing amazing trips we found the perfect equation for true friendship. Also, the pandemic experience would have been a totally different story without the climbing nights at Den Haag HS (always leaving with the 23.29 train) with Giorgio. I must confess our guilty obsessions for climbing was all I needed after a long day locked in my apartment. Lastly, I want to give a big hug to Hugo, because in our friendship I have literally found everything. Thank you for every time you smiled at me when I said it was 'the best day of my life', for the philosophical Spa sessions we enjoyed all around the Netherlands and for making me discover the conversational nature of reality, for our epic and emotional Classical concerts at the Concertgebouw, for sharing the passion for great food, wine and whisky, and for the amazing time spent with the Zoe community (Lea, Esther, Andreas, Filippo and Carolina).

I could never end this list without thanking my Girlfriend Dasha. We fired back to the long-distance relationship and a 2-year pandemic with crazy adventures across the world and many trains and flights to and from the Netherlands. Being on your side, I feel like I have suddenly become many Lorenzos at once. In sparse order: an expert in architecture, a connoisseur of healthy recipes, a great homemaker, a HIT trainer, and a true Londoner.

Then, to my close family (Mamma, Papà and Marco), my relatives and to Pierandrea I want to say: I have been so distant (maybe too distant) from you during these years, but I will never forget how much credit you deserve for shaping me the way I am. You truly are the reason why I will always miss Venice so much.

Finally, I want to express my deepest gratitude to my grandfather, nonno Nino. When I was struggling with my childhood, he was the one that stood up and fought with me. The steepest part of my personal growth was triggered by his love for Classical music and his genuine caring for my future. To him I want to direct my deepest gratitude.

# BIBLIOGRAPHY


Aglietti, G., Taylor, B., Fellowes, S., Salmon, T., Retat, I., Hall, A., Chabot, T., Pisseloup, A., Cox, C., Zarkesh, A., Mafficini, A., Vinkoff, N., Bashford, K., Bernal, C., Chaumette, F., Pollini, A., & W.H., S. (2020). The active space debris removal mission removeDebris. part 2: In orbit operations. *Acta Astronautica, 168*, 310–322. https://doi.org/10.1016/j.actaastro.2019.09.001

Al-Isawi, M., & Sasiadek, J. (2018). Guidance and control of a robot capturing an uncooperative space target. *Journal of Intelligent & Robotic Systems*, 1–9. https://doi.org/10.1007/s10846-018-0874-9

Allende-Alba, G., D'Amico, S., & Montenbruck, O. (2009). Radio frequency sensor fusion for relative navigation of formation flying satellites. *International Journal of Space Science and Engineering, 3*(2), 129–147. https://doi.org/10.1504/IJSPACESE.2015.072333

Barad, K. (2020). *Robust Navigation Framework for Proximity Operations around Uncooperative Spacecraft* [MSc Thesis]. Delft University of Technology.

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding, 110*(3), 346–359. https://doi.org/10.1016/j.cviu.2007.09.014

Bechini, M., Civardi, G., Quirino, M., Colombo, A., & Lavagna, M. Robust monocular pose initialization via visual and thermal image fusion. In: *73rd international astronautical congress*. Paris, Fance, 2022.

Ben-David, S., Blitzer, J., Crammer, K., & Pereira, F. (2007). Analysis of representations for domain adaptation. In B. Scholkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems*. MIT Press.

Benninghoff, H., & Boge, T. Rendezvous Involving a Non-cooperative, Tumbling Target - Estimation of Moments of Inertia and Center of Mass of an Unknown Target. In: *International symposium on space flight dynamics. 25*. Munich, Germany, 2015.

Biesbroek, R., Aziz, S., Wolahan, A., Cipolla, S., Richard-Noca, M., & Pigue, L. The Clearspace-1 mission: ESA and Clearspace team up to remove debris. In: *8th European conference on space debris*. Darmstadt, Germany, 2021.

Black, K., Shankar, S., Fonseka, D., Deutsch, J., Dhir, A., & Akella, M. Real-time, flight-ready, non-cooperative spacecraft pose estimation using monocular imagery. In: *31st AAS/AIAA space flight mechanics meeting*. Virtual, 2021.

Black, K., Shankar, S., Fonseka, D., Deutsch, J., Dhir, A., & Akella, M. Real-time, flight-ready, non-cooperative spacecraft pose estimation using monocular imagery. In: *31st AAS/AIAA space flight mechanics meeting*. Charlotte, NC, USA, 2021. https://doi.org/10.48550/arXiv.2101.09553.

Black, K., Shankar, S., Fonseka, D., Deutsch, J., Dhir, A., & Akella, M. Real-time, flight-ready, non-cooperative spacecraft pose estimation using monocular imagery. In: *31st AAS/AIAA space flight mechanics meeting*. Virtual, 2021.

Blackerby, C., Okamoto, A., Iizuka, S., Kobayashi, Y., Fujimoto, K., Seto, Y., Fujita, S., Iwai, T., Okada, N., Foreshaw, J., & Bradford, A. The Elsa-d end-of-life debris removal mission: Preparing for launch. In: *70th international astronautical congress*. Washington, DC, USA, 2019.

Boley, A., & Byers, M. (2021). Satellite mega-constellations create risks in Low Earth Orbit, the atmosphere and on Earth. *Scientific Report, 11*(10642). https://doi.org/10.1038/s41598-021-89909-7

Branco, J., Barrena, V., Escorial Olmos, D., Tarabini-Castellani, L., & Cropp, A. (2015). The formation flying navigation system for Proba 3. *Annual Review of Earth and Planetary Sciences, 24*, 37–47. https://doi.org/10.1007/978-3-319-13853-4-4

Calonder, M., Lepetit, V., Strecha, C., & Fua, P. BRIEF: Binary Robust Independent Elementary Features. In: *European conference on computer vision*. 2010, 778–792. https://doi.org/10.1007$/$978-3-642-15561-1$_$56.

Capuano, V., Alimo, S., Ho, A., & Chung, S. Robust Features Extraction for On-board Monocular-based Spacecraft Pose Acquisition. In: *AIAA Scitech forum*. San Diego, CA, USA, 2019. https://doi.org/10.2514/6.2019-2005.

Capuano, V., Kim, K., Harvard, A., & Chung, S. (2020). Monocular-based pose determination of uncooperative space objects. *Acta Astronautica, 166*, 493–506. https://doi.org/10.1016/j.actaastro.2019.09.027

Cavenago, F., Massari, M., Servadio, S., & Wittig, A. DA-based nonlinear filters for spacecraft relative state estimation. In: *2018 space flight mechanics meeting*. Kissimmee, FL, USA, 2018. https://doi.org/10.2514/6.2018-1964.

Cavrois, B., Vergnol, A., Donnard, A., Casiez, P., Southivong, U., Mongrard, O., Ankersen, F., Pezant, C., Breteker, P., Kolb, F., & Windmüller, M. (2015). LIRIS demonstrator on ATV5: A step beyond for European non cooperative navigation system. *AIAA Guidance, Navigation and Control Conference*. https://doi.org/10.2514/6.2015-0336

Chen, B., Cao, J., Parra, A., & Chin, T. Satellite Pose Estimation with Deep Landmark Regression and Nonlinear Pose Refinement. In: *International conference on computer vision*. Seoul, South Korea, 2019.

Colmenarejo, P., Binet, G., Strippoli, L., Peters, T., & Graziano, M. GNC aspects for active debris removal. In: *Proceedings of the eurognc 2013, 2nd ceas specialist conference on guidance, navigation & control*. 2013.

Colmenarejo, P., Graziano, M., Novelli, G., Mora, D., Serra, P., Tomassini, A., Seweryn, K., Prisco, G., & Gil Fernandez, J. (2019). On ground validation of debris removal technologies. *Acta Astronautica, 158*, 206–2019. https://doi.org/10.1016/j.actaastro.2018.01.026

Colombo, A., Civardi, G., Bechini, M., Quirino, M., & Lavagna, M. VIS-TIR cameras data fusion to enhance relative navigation during in orbit servicing operations. In: *73rd international astronautical congress*. Paris, Fance, 2022.

Crassidis, J., & Markley, F. (2012). Unscented filtering for spacecraft attitude estimation. *Journal of Guidance, Control, and Dynamics, 26*(4), 536–542. https://doi.org/10.2514/2.5102

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. (2018). Generative Adversarial Networks: An overview. *IEEE Signal Processing Magazine, 35*, 53–65. https://doi.org/10.1109/MSP.2017.2765202

Cui, J., Min, C., Bai, X., & Cui, J. (2019). An improved pose estimation method based on projection vector with noise error uncertainty. *IEEE Photonics Journal, 11*(2). https://doi.org/10.1109/JPHOT.2019.2901811

Curtis, H. (2005). *Orbital mechanics for engineering students*. Elsevier.

Dai, J., Li, Y., He, K., & Sun, J. Object detection via region-based fully convolutional networks. In: *Arxiv preprint*. 2016. https://doi.org/arXiv:1605.06409.

D'Amico, S. (2010). *Autonomous formation flying in low Earth orbit* (Doctoral dissertation). Delft University of Technology.

D'Amico, S. et al. Prisma. In: *Distributed space missions for earth system monitoring*. 2013, 599–637. https://doi.org/10.1007/978-1-4614-4541-8.

D'Amico, S., Benn, M., & Jorgensen, J. (2014). Pose estimation of an uncooperative spacecraft from actual space imagery. *International Journal of Space Science and Engineering, 2*(2), 171–189. https://doi.org/10.1504/IJSPACESE.2014.060600

David, P., DeMenthon, D., Duraiswami, R., & Samet, H. (2004). SoftPOSIT: Simultaneous pose and correspondence determination. *International Journal of Computer Vision, 59*(3), 259–284. https://doi.org/10.1023/B:VISI.0000025800.10423.1f

Davis, J., & Pernicka, H. (2019). Proximity operations about and identification of non-cooperative resident space objects using stereo imaging. *Acta Astronautica, 155*, 418–425. https://doi.org/10.1016/j.actaastro.2018.10.033

De Maesschalck, R., Jouan-Rimbaud, D., & Massart, L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems, 50*(1), 1–18. https://doi.org/10.1016/S0169-7439(99)00047-7

Deloo, J., & Mooij, E. (2015). Active debris removal : Aspects of trajectories, communication and illumination during final approach. *Acta Astronautica, 117*, 277–295. https://doi.org/10.1016/j.actaastro.2015.08.001

Dementhon, D., & Davis, L. (1995). Model-based object pose in 25 lines of code. *International Journal of Computer Vision, 15*(1&2), 123–141. https://doi.org/10.1007/BF01450852

Donahue, J., Krahenbuhl, P., & Darrell, T. Adversarial feature learning. In: *Int. conf. learning representation*. Toulon, France, 2017.

Drummond, T., & Cipolla, R. (2002). Real-time visual tracking of complex structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(7), 932–946. https://doi.org/10.1109/TPAMI.2002.1017620

Du, X., Liang, B., Xu, W., & Qiu, Y. (2011). Pose measurement of large non-cooperative satellite based on collaborative cameras. *Acta Astronautica, 68*(11&12), 2047–2065. https://doi.org/10.1016/j.actaastro.2010.10.021

Dubanchet, V., Bejar Romero, J., Gregertsen, K., Austad, H., Gancet, J., Natusiewicz, K., Vinals, J., Guerra, G., Rekleitis, G., Paraskevas, I., Nanos, K., Papadopoulos, E., Majewski, L., Ferraris, S., Purnell, J., Casu, D., D'Amico, J., & Andiappane, S. EROSS project – European autonomous robotic vehicle for on-orbit servicing. In: *I-sairas virtual conference*. Virtual, 2020.

Duda, R., & Hart, P. (1972). Use of the Hough Transformation to detect lines and curves in pictures. *Communications of the ACM, 15*(1), 11–15. https://doi.org/10.1145/361237.361242

Dutta, A., & Tsiotras, P. (2009). Hohmann-Hohmann and Hohmann-Phasing cooperative rendezvous maneuvers. *The Journal of the Astronautical Sciences, 57*, 393–417. https://doi.org/10.1007/BF03321510

Fehse, W. (2003). *Automated rendezvous and docking of spacecraft, 1st ed.* Cambridge University Press.

Felicetti, L., Sabatini, M., Pisculli, A., Gasbarri, P., & Palmerini, G. (2014). Adaptive thrust vector control during On-Orbit Servicing. *AIAA SPACE 2014 Conference and Exposition*. https://doi.org/10.2514/6.2014-4341

Ferraz, L., Binefa, X., & Moreno-Noguer, F. Leveraging Feature Uncertainty in the PnP Problem. In: *Proceedings of the british machine vision conference*. Nottingham, UK, 2014. https://doi.org/10.5244/C.28.83.

Ferraz, L., Binefa, X., & Moreno-Noguer, F. Very Fast Solution to the PnP Problem with Algebraic Outlier Rejection. In: *IEEE conference on computer vision and pattern recognition*. Columbus, OH, USA, 2014. https://doi.org/10.1109/CVPR.2014.71.

Filipe, N., Kontitsis, M., & Tsiotras, P. Extended Kalman Filter for Spacecraft Pose Estimation Using Dual Quaternions. In: *2015 American control conference*. Chicago, IL, USA, 2015, 3187–3192. https://doi.org/10.1109/ACC.2015.7171823.

Fischer, M. A., & Bolles, R. (1981). Random Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM, 24*(6), 381–395. https://doi.org/10.1145/358669.358692

Fischler, M., & Bolles, R. (1980). Random sample consensus: A paradigm for model fitting with applications to imageanalysis and automated cartography. *Communications of the Acm, 24*(6), 381–395. https://doi.org/10.1145/358669.358692

Flores-Abad, A., Ma, O., Pham, K., & Ulrich, S. (2014). A review of space robotics technologies for on-orbit servicing. *Progress in Aerospace Sciences, 68*, 1–26. https://doi.org/10.1016/j.paerosci.2014.03.002

Forshaw, J., Aglietti, G., Fellowes, S., Salmon, T., Retat, I., Hall, A., Chabot, T., Pisseloup, A., Tye, D., Bernal, C., Chaumette, F., Pollini, A., & W.H., S. (2020). The active space debris removal mission removeDebris. part 1: From concept to launch. *Acta Astronautica, 168*, 293–309. https://doi.org/10.1016/j.actaastro.2019.09.002

Forshaw, J., Aglietti, G., Navarathinam, N., Kadhem, H., Salmon, T., Pisseloup, A., Joffre, E., Chabot, T., Retat, I., Axthelm, T., Barraclough, S., Ratcliffe, A., Bernal, C., Chaumette, F., Pollini, A., & W.H., S. (2016). RemoveDebris: An in-orbit active debris removal demonstration mission. *Acta Astronautica, 127*, 448–463. https://doi.org/10.1016/j.actaastro.2016.06.018

Fraser, C. T., & Ulrich, S. (2021). Adaptive extended Kalman filtering strategies for spacecraft formation relative navigation. *Acta Astronautica, 178*. https://doi.org/10.1016/j.actaastro.2020.10.016

Gaias, G., & Ardaens, J.-S. (2018). In-orbit experience and lessons learned from the AVANTI experiment. *Acta Astronautica, 153*, 383–393. https://doi.org/10.1016/j.actaastro.2018.01.042

Galante, J., Van Eepoel, J., D' Souza, C., & Patrick, B. Fast Kalman Filtering for Relative Spacecraft Position and Attitude Estimation for the Raven ISS Hosted Payload. In: *39th AAS guidance and control conference*. Breckenridge, CO, USA, 2016.

Gansmann, M., Mongrard, O., & Ankersen, F. 3D Model-Based Relative Pose Estimation for Rendezvous and Docking Using Edge Features. In: Salzburg, Austria, 2017.

Gasbarri, P., Sabatini, M., & Palmerini, G. (2014). Ground tests for vision based determination and control of formation flying spacecraft trajectories. *Acta Astronautica*, *102*, 378–391. https://doi.org/10.1016/j.actaastro.2013.11.035

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., F.A., W., & Brendel, W. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. In: *International conference on learning representations*. New Orleans, LA, USA, 2019.

Ghifary, M., Kleijn, W., Zhang, M., & Balduzzi, D. Deep reconstruction-classification networks for unsupervised domain adaptation. In: *Int. conf. computer vision*. Rome, Italy, 2016.

Gill, E., & Montenbruck, O. (2012). *Satellite orbits - models, methods and applications*. Springer.

Giuffrida, G., Fanucci, L., Meoni, G., Batic, M., Buckley, L., Dunne, A., van Dijk, C., Esposito, M., Hefele, J., Vercruyssen, N., Furano, G., Pastena, M., & Aschbacher, J. (2021). The Φ-Sat-1 mission: The first on-board deep neural network demonstrator for satellite earth observation. *IEEE Transactions on Geoscience and Remote Sensing*, *60*. https://doi.org/10.1109/TGRS.2021.3125567

Golda, T., Kalb, T., Schumann, A., & Beyerer, J. Human pose estimation for real-world crowded scenarios. In: *6th IEEE international conference on advanced video and signal based surveillance (avss)*. Taipei, Taiwan, 2019, 1–8. https://doi.org/10.1109/AVSS.2019.8909823.

Guffanti, T., D'Amico, S., & Lavagna, M. (2017). Long term analytical Propagation of satellite relative motion in perturbed orbits. *Advances in the Astronautical Sciences Spaceflight Mechanics*, *160*, 355.

Hamel, J., & de Lafontaine, J. (2007). Linearized dynamics of formation flying spacecraft on a J2-perturbed elliptical orbit. *Journal of Guidance, Control, and Dynamics*, *30*(6), 1649–1658. https://doi.org/10.2514/1.29438

Hannah, S. ULTOR passive pose and position engine for spacecraft relative navigation. In: *Spie proceedings, sensors and systems for space applications ii. 6958*. 2008. https://doi.org/10.1117/12.777193.

Harvard, A., Capuano, V., Shao, E., & Chung, S.-J. Spacecraft Pose Estimation from Monocular Images Using Neural Network Based Keypoints and Visibility Maps. In: *AIAA Scitech 2020 forum*. Orlando, FL, USA, 2020. https://doi.org/10.1109/AERO.2018.8396425.

He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. In: *2016 IEEE conference on computer vision and pattern recognition*. 2016, 770–778.

Hou, X., Ma, C., Wang, Z., & Yuan, J. (2017). Adaptive pose and inertial parameters estimation of free-floating tumbling space objects using dual vector quaternions. *Advances in Mechanical Engineering*, *9*(10), 1–17. https://doi.org/10.1177/1687814017714210

Hou-Yuan, L., & Chang-Yin, Z. (2018). An estimation of Envisat's rotational state accounting for the precession of its rotational axis caused by gravity-gradient torque. *Advances in Space Research, 61*(1), 182–188. https://doi.org/10.1016/j.asr.2017.10.014

Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. In: *Arxiv preprint*. 2017. https://doi.org/arXiv:1704.04861.

Huo, Y., Li, Z., & Zhang, F. (2020). Fast and accurate spacecraft pose estimation from single shot space imagery using box reliability and keypoints existence judgments. *IEEE Access, 8*. https://doi.org/10.1109/ACCESS.2020.3041415

Jackson, P., A., A.-A., Bonner, S., Breckon, T., & Obara, B. Style augmentation: Data augmentation via style randomization. In: *Conference on computer vision and pattern recognition*. Salt Lake City, UT, USA, 2018.

Julier, S., & Uhlmann, J. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE, 92*(3), 401–422. https://doi.org/10.1109/JPROC.2003.823141

Kailath, T. (1968). An innovations approach to least-squares estimation–Part I: Linear filtering in additive white noise. *IEEE Transactions on Automatic Control, 13*(6). https://doi.org/10.1109/TAC.1968.1099025

Karlgaard, C. (2010). *Robust adaptive estimation for autonomous rendezvous in elliptical orbit* (Doctoral dissertation). Virginia Polytechnic Institute and State University.

Kessler, D., & Cour-Palais, B. (1978). Collision frequency of artificial satellites: The creation of a debris belt. *Journal of Geophysical Research, 83*(6). https://doi.org/10.1029/JA083iA06p02637

Kim, S., Crassidis, J., Cheng, Y., & Fosbury, A. (2007). Kalman filtering for relative spacecraft attitude and position estimation. *Journal of Guidance, Control, and Dynamics, 30*(1), 133–143. https://doi.org/10.2514/1.22377

Kingma, D., & Ba, J. Adam: A method for stochastic optimization. In: *3rd international conference for learning representations*. San Diego, CA, USA, 2015. https://doi.org/10.2514/6.2018-2100.

Kisantal, M., Sharma, S., Park, T., Izzo, D., Martens, M., & D'Amico, S. (2020). Satellite Pose Estimation Challenge: Dataset, competition design and results. *IEEE Transactions on Aerospace and Electronic Systems*. https://doi.org/10.1109/TAES.2020.2989063

Koenig, A., Guffanti, T., & D'Amico, S. (2017). New state transition matrices for spacecraft relative motion in perturbed orbits. *Journal of Guidance, Control, and Dynamics, 40*(7), 1749–1768. https://doi.org/10.2514/1.G002409

Kozlowsky, L., & Kosonocky, W. (1995). Handbook of optics. McGraw-Hill.

Kozlowsky, L., & Kosonocky, W. (2010). Yearbook on space policy 2008/2009. Springer.

Krizhevsky, A., Sutskever, I., & Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. In: *26th annual conference on neural information processing systems. 1*. Lake Tahoe, NV, USA, 2012, 1097–1105.

Krüger, H., & Theil, S. TRON - hardware-in-the-loop test facility for lunar descent and landing optical navigation. In: *IFAC-ACA automatic control in aerospace*. Nara, Japan, 2010.

Lefferts, E., Markley, F., & Shuster, M. (1982). Kalman filtering for spacecraft attitude estimation. *Journal of Guidance, Control, and Dynamics*, *5*(5), 417–429. https://doi.org/10.2514/3.56190

Lepetit, Moreno-Noguer, F., & Fua, P. (2009). EPnP: An accurate O(n) solution to the PnP problem. *International Journal of Computer Vision*, *81*, 155–166. https://doi.org/10.1007/s11263-008-0152-6

Lepetit, V., & Fua, P. (2005). Monocular model-based 3D tracking of rigid objects: A survey. *Foundations and Trends in Computer Graphics and Vision*, *1*(1), 1–89. https://doi.org/10.1561/0600000001

Liu, C., & Hu, W. (2014). Relative pose estimation for cylinder-shaped spacecrafts using single image. *IEEE Transactions on Aerospace and Electronics Systems*, *50*(4), 3036–3056. https://doi.org/10.1109/TAES.2014.120757

Long, A., Richards, M., & Hastings, D. (2007). On-orbit servicing: A new value proposition for satellite design and operation. *Journal of Spacecraft and Rockets*, *44*(4), 964–976.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Mahendran, S., Ali, H., & Vidal, R. 3d pose regression using convolutional neural networks. In: *Arxiv preprint*. 2017. https://doi.org/10.48550/arXiv.1708.05628.

Markley, F. (2003). Attitude error representations for Kalman filtering. *Journal of Guidance, Control, and Dynamics*, *26*(2), 311–317. https://doi.org/10.2514/2.5048

McDowell, J. (2020). The Low Earth Orbit satellite population and impacts of the SpaceX starlink constellation. *The Astrophysical Journal Letters*, *892*(2). https://doi.org/10.3847/2041-8213/ab8016

Mehra, R. (1972). Approaches to adaptive filtering. *IEEE Transactions on Automatic Control*, *17*(5), 693–698. https://doi.org/10.1109/TAC.1972.1100100

Meng, Q., Liang, J., & Ma, O. Estimate of All the Inertial Parameters of a Free-Floating Object in Orbit. In: *2018 AIAA guidance, navigation and control conference*. Kissimmee, FL, USA, 2018. https://doi.org/10.2514/6.2018-1606.

Merriaux, P., Dupuis, Y., Boutteau, R., Vasseur, P., & Savatier, X. (2017). A study of Vicon system positioning performance. *Sensors*, *17*(7), 1591. https://doi.org/10.3390/s17071591

Mitchell, I. (2011). Draper laboratory overview of rendezvous and capture operations [R/OL].

Mohamed, A. H., & Schwarz, K. P. (1999). Adaptive Kalman filtering for INS/GPS. *Journal of Geodesy*, *73*(4), 193–203. https://doi.org/10.1007/s001900050236

Mortensen, R. (1968). Maximum-likelihood recursive nonlinear filtering. *Journal of Optimization Theory and Applications*, *2*(6), 386–394. https://doi.org/10.1007/BF00925744

Myers, K., & Tapley, B. (1976). Adaptive sequential estimation with unknown noise statistics. *IEEE Transactions on Automatic Control*, *21*, 520–523. https://doi.org/10.1109/TAC.1976.1101260

Myers, K. A. (1974). *Filtering theory methods and applications to the orbit determination problem for near-Earth satellites* (Doctoral dissertation). The University of Texas at Austin.

Naasz, B., Burns, R., Queen, S., Van Eepoe, J., Hannah, J., & Skelton, E. (2009). The HST SM4 relative navigation sensor system: Overview and preliminary testing results from the flight robotics lab. *The Journal of the Astronautical Sciences*, *57*(1&2), 457–483. https://doi.org/10.1007/BF03321512

Newell, A., Yang, K., & Deng, J. (2016). Stacked hourglass networks for human pose estimation. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer vision - eccv 2016* (pp. 483–499). Springer.

Oberkampf, D., Dementhon, D., & Davis, L. (1996). Iterative pose estimation using coplanar feature points. *Computer Vision and Image Understanding*, *63*(3), 495–511. https://doi.org/10.1006/cviu.1996.0037

Opromolla, R., Fasano, G., Rufino, G., & Grassi, M. (2015). Uncooperative pose estimation with a LIDAR-based system. *Acta Astronautica*, *110*, 287–297. https://doi.org/10.1016/j.actaastro.2014.11.003

Opromolla, R., Fasano, G., Rufino, G., & Grassi, M. (2017a). Pose estimation for spacecraft relative navigation using model-based algorithms. *IEEE Transactions On Aerospace And Electronic Systems*, *53*, 431–447. https://doi.org/10.1109/TAES.2017.2650785

Opromolla, R., Fasano, G., Rufino, G., & Grassi, M. (2017b). A review of cooperative and uncooperative spacecraft pose determination techniques for close-proximity operations. *Progress in Aerospace Sciences*, *93*, 53–72. https://doi.org/10.1016/j.paerosci.2017.07.001

Ostrowsky, A. (1966). *Solution of equations and systems of equations* (2nd ed.). Academic Press, New York.

Pardini, C., & Anselmo, L. (2020). Environmental sustainability of large satellite constellations in Low Earth Orbit. *Acta Astronautica*, *170*, 27–36. https://doi.org/10.1016/j.actaastro.2020.01.016

Park, T., Bosse, J., & D'Amico, S. Robotic testbed for rendezvous and optical navigation: Multi-source calibration and machine learning use cases. In: *AAS/AIAA astrodynamics specialist conference*. Big Sky, MT, USA, 2021.

Park, T., & D'Amico, S. Adaptive neural network-based unscented Kalman filter for spacecraft pose tracking at rendezvous. In: *AAS/AIAA astrodynamics specialist conference*. Charlotte, NC, USA, 2022.

Park, T., & D'Amico, S. Robust multi-task learning and online refinement for spacecraft pose estimation across domain gap. In: *11th international workshop on satellite constellations & formation flying*. Milan, ITA, 2022.

Park, T., Martens, M., Lecuyer, G., Izzo, D., & D'Amico, S. Speed+: Next-generation dataset for spacecraft pose estimation across domain gap. In: *Arxiv preprint*. 2021. https://doi.org/10.48550/arXiv.2110.03101.

Park, T., Sharma, S., & D'Amico, S. Towards Robust Learning-Based Pose Estimation of Noncooperative Spacecraft. In: *AAS/AIAA astrodynamics specialist conference*. Portland, ME, USA, 2019. https://doi.org/10.48550/arXiv.1909.00392.

Pasqualetto Cassinis, L. et al. (2022a). On-ground validation of a CNN-based monocular pose estimation system for uncooperative spacecraft: Bridging domain shift in rendezvous scenarios. *Acta Astronautica, 196*, 123–138. https://doi.org/10.1016/j.actaastro.2022.04.002

Pasqualetto Cassinis, L., Fonod, R., & Gill, E. (2019). Review of the robustness and applicability of monocular pose estimation systems for relative navigation with an uncooperative spacecraft. *Progress in Aerospace Sciences, 110*. https://doi.org/10.1016/j.paerosci.2019.05.008

Pasqualetto Cassinis, L., Fonod, R., Gill, E., Ahrns, I., & Gil Fernandez, J. CNN-Based Pose Estimation System for Close-Proximity Operations Around Uncooperative Spacecraft. In: *AIAA Scitech 2019 forum*. Orlando, FL, USA, 2020. https://doi.org/10.2514/6.2020-1457.

Pasqualetto Cassinis, L., Fonod, R., Gill, E., Ahrns, I., & Gil-Fernandez, J. (2021). Evaluation of tightly- and loosely-coupled approaches in CNN-based pose estimation systems for uncooperative spacecraft. *Acta Astronautica, 182*, 189–202. https://doi.org/10.1016/j.actaastro.2021.01.035

Pasqualetto Cassinis, L., Menicucci, A., Gill, E., Ahrns, I., & Gil Fernandez, J. On-ground validation of a cnn-based monocular pose estimation system for uncooperative spacecraft. In: *8th European conference on space debris*. Darmstadt, Germany, 2021.

Pasqualetto Cassinis, L., Park, J., Stacey, N., D'Amico, S., Menicucci, A., Gill, E., Ahrns, I., & Sanchez-Gestido, M. (2022b). Leveraging neural network uncertainty in adaptive unscented Kalman filter for spacecraft pose estimation. *Submitted Manuscript*.

Pavlakos, G., Zhou, X., Chan, A., Derpanis, K., & Daniilidis, K. 6-DoF Object Pose from Semantic Keypoints. In: *IEEE international conference on robotics and automation*. Marina Bay Sands, Singapore, 2017.

Pesce, V., Haydar, M., Lavagna, M., & Lovera, M. (2019). Comparison of filtering techniques for relative attitude estimation of uncooperative space objects. *Aerospace Science and Technology, 84*, 318–328. https://doi.org/10.1016/j.ast.2018.10.031

Pesce, V., Lavagna, M., & Bevilacqua, R. (2017). Stereovision-based pose and inertia estimation of unknown and uncooperative space objects. *Advances in Space Research, 59*, 236–251. https://doi.org/10.1016/j.asr.2016.10.002

Pesce, V., Opromolla, R., Sarno, S., Lavagna, M., & Grassi, M. (2019). Autonomous relative navigation around uncooperative spacecraft based on a single camera. *Aerospace Science and Technology, 84*, 1070–1080. https://doi.org/10.1016/j.ast.2018.11.042

Picone, J., Hedin, A., Drob, D., & Aikin, A. (2002). NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. *Journal of Geophysical Research: Space Physics, 107*(A12), SIA 15–1–SIA 15–16. https://doi.org/10.1029/2002JA009430

Pyrak, M., & Anderson, J. Performance of Northrop Grumman's mission extension vehicle (mev) RPO imagers at GEO. In: *Proceedings of autonomous systems: Sensors, processing and security for ground, air, sea and space vehicles and infrastructure. 12115*. Orlando, FL, USA, 2022.

Qiu, S., Guo, Y., Xing, J., & Ma, G. Inertia Parameter and Attitude Estimation of Space Noncooperative Tumbling Target Based on a Two-Step Method. In: *Iecon 2017-*

*43rd annual conference of the IEEE industrial electronics society*. Beijing, China, 2017. https://doi.org/10.1109/IECON.2017.8217143.

Reintsema, D., Thaeter, J., Rathke, A., Naumann, W., Rank, P., & Sommer, J. Deos – the german robotics approach to secure and de-orbit malfunctioned satellites from low earth orbits. In: *International symposium on artificial intelligence, robotics and automation in space*. Sapporo, Japan, 2010.

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(6), 1137 –1149. https://doi.org/10.1109/TPAMI.2016.2577031

Ries, J., Bettadpur, S., Eanes, R. J., Kang, Z., Ko, U.-D., McCullough, C., Nagel, P., Pie, N., Poole, S., Richter, T., Save, H., & Tapley, B. (2016). *The development and evaluation of the global gravity model ggm05* (tech. rep. CSR-16-02). Center for Space Research, The University of Texas at Austin.

Rondao, D., & Aouf, N. Multi-View Monocular Pose Estimation for Spacecraft Relative Navigation. In: *2018 AIAA guidance, navigation, and control conference*. Kissimmee, FL, USA, 2018. https://doi.org/10.2514/6.2018-2100.

Rondao, D., Aouf, N., & Dubois-Matra, O. Multispectral Image Processing for Navigation Using Low Performance Computing. In: *69th international astronautical congress*. Bremen, Germany, 2018.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention*, 234–241.

Sagnières, L., & Sharf, I. (2019). Long-term rotational motion analysis and comparison to observations of the inoperative Envisat. *Journal of Guidance, Control, and Dynamics, 42*(2). https://doi.org/10.2514/1.G003647

Sakkos, D., Shum, H., & Ho, E. Illumination-based data augmentation for robust background subtraction. In: *Proceedings of the 2019 international conference on software knowledge information management and applications (skima)*. 2019.

Schmidt, S. (1966). Applications of state space methods to navigation problems. *Advances in Control Systems, 3*, 293–340. https://doi.org/10.1016/B978-1-4831-6716-9.50011-4

Schnitzer, F., Sonnenburg, A., Janschek, K., & Sanchez Gestido, M. Lessons-learned from On-ground Testing of Image-based Non-cooperative Rendezvous Navigation with Visible-spectrum and Thermal Infrared Cameras. In: *10th international esa conference on guidance, navigation, and control systems*. Salzburg, Austria, 2017.

Segal, S., Gurfil, P., & Shahid, K. (2014). In-orbit tracking of resident space objects: A comparison of monocular and stereoscopic vision. *IEEE Transactions on Aerospace and Electronic Systems, 50*(1), 676–688. https://doi.org/10.1109/TAES.2013.120006

Setterfield, T., Miller, D., Saenz Otero, A., Frazzoli, E., & Leonard, J. (2018). Inertial properties estimation of a passive On-Orbit object using Polhode analysis. *Journal of Guidance, Control, and Dynamics, 41*(10), 2214–2231. https://doi.org/10.2514/1.G003394

Sharma, S., Beierle, C., & D'Amico, S. Pose Estimation for Non-Cooperative Spacecraft Rendezvous using Convolutional Neural Networks. In: *IEEE aerospace conference*. Big Sky, MT, USA, 2018. https://doi.org/10.1109/AERO.2018.8396425.

Sharma, S., & D'Amico, S. (2015). Comparative Assessment of techniques for initial pose estimation using monocular vision. *Acta Astronautica, 123*, 435–445. https://doi.org/10.1016/j.actaastro.2015.12.032

Sharma, S., & D'Amico, S. Reduced-Dynamics Pose Estimation for Non-Cooperative Spacecraft Rendezvous Using Monocular Vision. In: *38th AAS guidance and control conference*. Breckenridge, CO, USA, 2017.

Sharma, S., & D'Amico, S. Pose Estimation for Non-Cooperative Spacecraft Rendezvous using Neural Networks. In: *29th AAS/AIAA space flight mechanics meeting*. Ka'anapali, HI, USA, 2019. https://doi.org/10.1109/AERO.2018.8396425.

Sharma, S., Ventura, J., & D'Amico, S. (2018). Robust model-based monocular pose initialization for noncooperative spacecraft rendezvous. *Journal of Spacecraft and Rockets, 55*(6), 1–16. https://doi.org/10.2514/1.A34124

Sheinfeld, D., & Rock, S. (2009). Rigid body inertia estimation with applications to the capture of a tumbling satellite. *Advances in Astronautical Sciences, 134*, 343–356.

Shi, J., & Ulrich, S. (2021). Uncooperative spacecraft pose estimation using monocular monochromatic images. *Journal of Spacecraft and Rockets, 58*(2). https://doi.org/10.2514/1.A34775

Shi, J., Ulrich, S., & Ruel, S. Posenet: A convolutional network for real-time 6-dof camera relocalization. In: *IEEE international conference on computer vision (iccv)*. Santiago, CL, 2015. https://doi.org/10.1109/ICCV.2015.336.

Shi, J., Ulrich, S., & Ruel, S. Spacecraft pose estimation using a monocular camera. In: *67th international astronautical congress*. Guadalajara, Mexico, 2016.

Shi, J., Ulrich, S., & Ruel, S. Spacecraft Pose Estimation using Principal Component Analysis and a Monocular Camera. In: *AIAA guidance, navigation, and control conference*. Grapevine, TX, USA, 2017. https://doi.org/10.2514/6.2017-1034.

Shi, J., Ulrich, S., & Ruel, S. CubeSat Simulation and Detection using Monocular Camera Images and Convolutional Neural Networks. In: *2018 AIAA guidance, navigation, and control conference*. Kissimmee, FL, USA, 2018. https://doi.org/10.2514/6.2018-1604.

Shi, J., Ulrich, S., Ruel, S., & Anctil, M. Uncooperative spacecraft pose estimation using an infrared camera during proximity operations. In: *AIAA space 2015 conference and exposition*. Pasadena, CA, USA, 2015. https://doi.org/10.2514/6.2015-4429.

Simon, D. (2006). *State estimation: Kalman, $H_\infty$ and nonlinear approaches*. John Wiley & Sons.

Simonyan, K., & Zisserman, A. Very deep convolutional networks for large-scale image recognition. In: *Arxiv preprint*. 2014. https://doi.org/10.48550/arXiv.1409.1556.

Solà, J. (2017). Quaternion kinematics for the error-state Kalman filter. *arXiv e-prints*, Article arXiv:1711.02508, arXiv:1711.02508.

Sonawani, S., Alimo, R., Detry, R., Jeong, D., Hess, A., & Ben Amor, H. Assistive relative pose estimation for on-orbit assembly using convolutional neural networks. In: *AIAA Scitech 2020 forum*. Orlando, FL, USA, 2020. https://doi.org/10.1109/AERO.2018.8396425.

Stacey, N., & D'Amico, S. (2021). Adaptive and dynamically constrained process noise estimation for orbit determination. *IEEE Transactions on Aerospace and Electronic Systems*. https://doi.org/10.1109/TAES.2021.3074205

Stadnyk, K., Hovell, K., & Brewster, L. Space debris removal with sub-tethered net: A feasibility study and preliminary design. In: *8th European conference on space debris*. Darmstadt, Germany (Virtual), 2021.

Su, H., Qi, C., Li, Y., & Guibas, L. Render for cnn: Viewpoint estimation in images using cnns trained with rendered 3d model views. In: *Proceedings of the IEEE international conference on computer vision (iccv)*. 2015, 2686–2694. https://doi.org/10.48550/arXiv.1505.05641.

Sugiyama, M., & Muller, K.-R. (2005). Input-dependent estimation of generalization error under covariate shift. *Statistics & Decisions*, *23*(4), 249–279. https://doi.org/10.1524/stnd.2005.23.4.249

Sullivan, J., Grimberg, S., & D'Amico, S. (2017). Comprehensive survey and assessment of spacecraft relative motion dynamics models. *Journal of Guidance, Control, and Dynamics*, *40*(8), 1837–1859. https://doi.org/10.2514/1.G002309

Sullivan, J., & D'Amico, S. (2017). Nonlinear Kalman filtering for improved angles-only navigation using relative orbital elements. *Journal of Guidance, Control, and Dynamics*, *40*(9), 2183–2200. https://doi.org/10.2514/1.G002719

Sun, K., Xiao, B., Liu, D., & Wang, J. Deep high-resolution representation learning for human pose estimation. In: *2019 IEEE conference on computer vision and pattern recognition*. Long Beach, CA, USA, 2019.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Las Vegas, NV, USA, 2016, 2818–2826. https://doi.org/10.48550/arXiv.1512.00567.

Tabb, A., & Yousef, K. (2017). Solving the robot-world hand-eye (s) calibration problem with iterative methods. *Machine Vision and Applications*, 569–590. https://doi.org/10.1007/s00138-017-0841-7

Tatsch, A., Fitz-Coy, N., & Gladun, S. On-orbit Servicing: A brief survey. In: *Proceedings of the 2006 performance metrics for intelligent systems workshop*. 2006, 21–23.

Telaar, J., Ahrns, I., Estable, S., Rackl, W., De Stefano, M., Lampariello, R., Santos, N., Serra, P., Canetri, M., & Ankersen, F. GNC architecture for the e.Deorbit mission. In: *7th European conference for aeronautics and space sciences (eucass)*. 2017. https://doi.org/10.13009/EUCASS2017-317.

Thiele, S. Investigating the risks of debris-generating asat tests in the presence of mega-constellations. In: *Advanced maui optical and space surveillance technologies conference*. Maui, HI, USA, 2021.

Thrun, S., Burgard, W., & Fox, D. (2005). Probabilistic robotics. In R. C. Arkin (Ed.). The MIT Press.

Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., & Abbeel, P. Domain randomization for transferring deep neural networks from simulation to the real world. In: *Int. conf. intelligent robots and systems*. 2017, 23–30. https://doi.org/10.1109/IROS.2017.8202133.

Tweddle, B., & Saenz-Otero, A. (2015). Relative computer vision based navigation for small inspection spacecraft. *Journal of Guidance, Control, and Dynamics*, *38*(5), 969–978. https://doi.org/10.2514/1.G000687

Urban, S., Leitloff, J., & Hinz, S. (2016). MLPnP: A rel-time maximum likelihood solution to the perspective-n-point problem. *Proceedings of the British Machine Vision Conference*, *3*. https://doi.org/10.5194/isprsannals-III-3-131-2016

Valmorbida, A., Mazzucato, M., & Pertile, M. (2020). Calibration procedures of a vision-based system for relative motion estimation between satellites flying in proximity. *Measurement*, *151*. https://doi.org/10.1016/j.measurement.2019.107161

Wertz, J. (1978). Spacecraft attitude determination and control. In J. Wertz (Ed.). Reidel Dordrecht.

Wieser, M., Richard, H., Hausmann, G., Meyer, J.-C., Jaekel, S., Lavagna, M., & Biesbroek, R. E.deorbit mission: OHB debris removal concepts. In: *Astra 2015-13th symposium on advanced space technologies in robotics and automation*. Noordwijk, The Netherlands, 2015.

Wilde, M., Clark, C., & Romano, M. (2019). Historical survey of kinematic and dynamic spacecraft simulators for laboratory experimentation of on-orbit proximity maneuvers. *Progress in Aerospace Sciences*, *110*. https://doi.org/10.1016/j.paerosci.2019.100552

Xu, B., & Wang, S. Vision-Based Moment of Inertia Estimation of Non-Cooperative Space Object. In: *10th international symposium on computational intelligence and design*. Hangzhou, China, 2017. https://doi.org/10.1109/ISCID.2017.68.

Yamanaka, K., & Ankersen, F. (2002). New state transition matrix for relative motion on an arbitrary elliptical orbit. *Journal of Guidance, Control, and Dynamics*, *25*(1), 60–66. https://doi.org/10.2514/2.4875

Yilmaz, O. et al. Using infrared based relative navigation for active debris removal. In: *10th international esa conference on guidance, navigation, and control systems*. Salzburg, Austria, 2017.

Yilmaz, O., Aouf, N., Checa, E., Majewski, L., & Sanchez Gestido, M. 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*. SAGE Publications, 2017, 1–13. https://doi.org/10.1177/0954410017740917.

Yilmaz, O., Aouf, N., Majewski, L., & Sanchez Gestido, M. Evaluation of Feature Detectors for Infrared Imaging in View of Active Debris Removal. In: Darmstadt, Germany, 2017.

Zamani, M., Trumpf, J., & Mahony, R. (2013). Minimum-energy filtering for attitude estimation. *IEEE Transactions on Automatic Control*, *58*(11), 2917–2921. https://doi.org/10.1109/TAC.2013.2259092

Zamani, M., Trumpf, J., & Mahony, R. On the distance to optimality of the geometric approximate minimum-energy attitude filter. In: Portland, OR, USA, 2014, 4943–4948. https://doi.org/10.1109/ACC.2014.6858915.

Zhang, L., Li, T., Yang, H., Zhang, S., Cai, H., & Qian, S. (2015). Unscented Kalman filtering for relative spacecraft attitude and position estimation. *Journal of Navigation*, *68*(3), 528–548. https://doi.org/10.1017/S0373463314000769

Zwick, M., Huertas, I., Gerdes, L., & Ortega, G. ORGL - ESA's test facility for approach and contact operations in orbital and planetary environments. In: *International symposium on artificial intelligence, robotics and automation in space*. Madrid, Spain, 2018.

# CURRICULUM VITÆ

## LORENZO PASQUALETTO CASSINIS

15-09-1993 Born in Venice, Italy.

## EDUCATION

2018–2022 Ph.D. in Aerospace Engineering
Delft University of Technology, Delft, The Netherlands
*Thesis:* Monocular Vision-Based Pose Estimation of Uncooperative Spacecraft
*Promotor:* Prof. dr. E.K.A. Gill

2015–2017 M.Sc. in Aerospace Engineering
Delft University of Technology, Delft, The Netherlands
*Thesis:* Telemetry Analysis of Delfi-C$^3$

2012–2015 B.Sc. in Aerospace Engineering
Padua University, Padua, Italy

## MAIN RESEARCH EXPERIENCE

04/2022–05/2022 Visiting Researcher, Airbus Defence & Space
Bremen, Germany

09/2021–02/2022 Visiting Student Researcher, Stanford Rendezvous Laboratory
Stanford, CA, USA

02/2019–08/2021 Visiting Researcher, European Space Research and Technology Centre
Noordwijk, The Netherlands

11/2017–05/2018 Junior GNC Engineer, Gmv
Madrid, Spain

# List of Publications

## Journal Publications

4. **L. Pasqualetto Cassinis**, T.H. Park, N. Stacey, S. D'Amico, A. Menicucci, E. Gill, I. Ahrns, and M. Sanchez-Gestido, *Leveraging Neural Network Uncertainty in Adaptive Unscented Kalman Filter for Spacecraft Pose Estimation*, Advances in Space Research Journal - Submitted (2022).

3. **L. Pasqualetto Cassinis**, A. Menicucci, E. Gill, I. Ahrns, and M. Sanchez-Gestido, *On-Ground Validation of a CNN-based Monocular Pose Estimation System for Uncooperative Spacecraft: Bridging Domain Shift in Rendezvous Scenarios*, Acta Astronautica **196**, 123-138 (2021).

2. **L. Pasqualetto Cassinis**, R. Fonod, E. Gill, I. Ahrns, and J. Gil-Fernandez, *Evaluation of tightly- and loosely-coupled approaches in CNN-based pose estimation systems for uncooperative spacecraft*, Acta Astronautica **182**, 189-202 (2021).

1. **L. Pasqualetto Cassinis**, R. Fonod, and E. Gill, *Review of the Robustness and Applicability of Monocular Pose Estimation Systems for Relative Navigation with an Uncooperative Spacecraft*, Progress In Aerospace Sciences **110** (2019).

## Conference Publications

5. T. Hendrix, **L. Pasqualetto Cassinis**, D. Rijlaarsdam and L. Buckley, *Validating a CNN-based Pose Estimation System for Relative Navigation with an Uncooperative Spacecraft on Myriad X Space Grade Processor*, 2022 Clean Space Industry Days, Noordwijk, The Netherlands (2022).

4. **L. Pasqualetto Cassinis**, A. Menicucci, E. Gill, I. Ahrns, and M. Sanchez-Gestido, *Bridging The Domain Shift Of CNN-Based Pose Estimation Systems in Active Debris Removal Scenarios*, 73rd International Astronautical Congress, Paris, France (2022).

3. **L. Pasqualetto Cassinis**, A. Menicucci, E. Gill, I. Ahrns, and M. Sanchez-Gestido, *On-Ground Validation of a CNN-based Monocular Pose Estimation System for Uncooperative Spacecraft*, 8th European Conference on Space Debris, Darmstadt, Germany (2021).

2. **L. Pasqualetto Cassinis**, R. Fonod, E. Gill, I. Ahrns, and J. Gil-Fernandez, *CNN-Based Pose Estimation System for Close-Proximity Operations Around Uncooperative Spacecraft*, AIAA Scitech Forum, Orlando, FL, USA (2020).

1. **L. Pasqualetto Cassinis**, R. Fonod, E. Gill, I. Ahrns, and J. Gil-Fernandez, *Comparative Assessment of Image Processing Algorithms for the Pose Estimation of an Uncooperative Spacecraft*, International Workshop on Satellite Constellations & Formation Flying, Glasgow, UK (2019).

## BOOK CHAPTERS

1. S. Silvestrini, **L. Pasqualetto Cassinis**, R. Hinz, D. Gonzalez-Arjona, M. Tipaldi Sr., P. Visconti, F. Corradino, V. Pesce, and A. Colagrossi, *AI and Modern Applications - Modern Spacecraft GNC*, Modern Spacecraft Guidance, Navigation and Control: From System Modelling to AI and Innovative Applications, Editors: V. Pesce, A. Colagrossi, and S. Silvestrini, Elsevier Ltd. (2022).