UNIVERSITÀ
DEGLI STUDI
DI PADOVA

UNIVERSITA' DEGLI STUDI DI PADOVA

Dipartimento di Ingegneria Industriale DII

Corso di Laurea Magistrale in Ingegneria Aerospaziale

ANALYSIS AND VALIDATION OF A VISION-BASED POSE INITIALIZATION ALGORITHM FOR NON-COOPERATIVE SPACECRAFTS

Relatore: Prof. Stefano Debei

Correlatore: Prof. Marco Pertile

Pierdomenico Fracchiolla

Mat. 1153803

Anno Accademico 2018/2019

Table of Contents

Summary

This thesis addresses the problem of relative pose determination for on orbit-servicing and active debris removal missions in the context of non-cooperative passive spacecraft. The problem of relative attitude and position measurements is introduced and described, then, a survey of the state-of-the-art instruments and techniques employed in current and past space missions follows. Among the possible operative scenarios, the case of non-cooperative passive target vehicles is further explored in order to demonstrate the advantages of a monocular feature-based technique. The thesis introduces and discusses an architecture for model-based pose estimation, called SVD method, which adopts newer techniques to perform feature identification and matching. By outlining the subsystems of the SVD architecture, the main innovations and peculiarities of the method are explained and the results, obtained on real images of PRISMA mission's Tango spacecraft, are summarized. Subsequently, a version of the algorithm is implemented and validated with two different datasets. The SVD architecture is applied to a larger open dataset of synthetic imagery, thus, collecting more statistically relevant data over a variety of images with various illumination conditions and background interferences. Mean rotational and translational errors are evaluated, as well as the effectiveness of the edge detection techniques and the correlation between the success rate of pose determination, the identification of target's region of interest and the combination of geometric groups. A further validation test on real experimental images of a 2U-CubeSat mock-up is presented. Within the SPARTANS testbed for microsatellite relative dynamics, a dataset of images and respective pose reference values are obtained and fed to the SVD algorithm. The results show how the pose initialization success rate consistently improves at a shorter range on images without background interference and further confirm the validity of the assumption that coupling complex geometrical features and simple, uniquely identifiable linear features provides better chances of determining a correct pose. The validation process proves the limitations of the SVD architecture in terms of relative distances, spacecraft model characteristics, environmental conditions and overall feasibility over a more flexible operational range, yet it provides sufficiently accurate pose solutions to initialize navigation measurements during close proximity operations.

Riassunto Esteso

Il crescente bisogno di condurre operazioni a corto raggio in contesti spaziali ha spinto la ricerca aerospaziale verso lo sviluppo di sistemi autonomi di guida e navigazione. Numerose sono le tipologie di missione che prevedono l'avvicinamento e il rendezvous tra veicoli con diverse caratteristiche, tra cui missioni in formation flying (FF), missioni rivolte alla rimozione di satelliti non più operativi o detriti spaziali privi di controllo d'assetto, missioni di servizio per la manutenzione e il rifornimento di grandi spacecraft. In ognuno di questi casi, la capacità di un veicolo spaziale di integrare un sensore affidabile, veloce e preciso di misura della posa relativa, ossia orientazione e posizione, all'interno del proprio sistema di navigazione permette di ampliare notevolmente la quantità e la qualità della propria operatività.

Nella tesi si esplora la possibilità di impiegare una tecnica di misura della posa basata su un sistema di visione monoculare nello scenario di un satellite target non cooperativo e passivo. Per poter definire una tecnica di misura e un adeguato sensore, è fondamentale tener presente caratteristiche fondamentali del target come la cooperatività, cioè la capacità di instaurare una comunicazione diretta tra chaser e target, la passività, ossia la presenza di predefiniti elementi predisposti alla detezione come markers, LEDs, o Corner Cube Reflectors (CCRs) ed infine la conoscenza di informazioni a priori quali, ad esempio, la geometria e l'inerzia del satellite così come la tipologia di controllo d'assetto. Il primo capitolo della tesi è, perciò, incentrato sulla descrizione della varietà di scenari di missione in cui sono necessarie delle misure di posa relativa e un resoconto dello stato dell'arte dei sensori e delle tecniche impiegate per questo scopo.

Il secondo capitolo presenta la descrizione dettagliata di un metodo per l'inizializzazione delle misure di posa basato sulle acquisizioni di un sistema di visione monoculare. Questa tecnica, denominata metodo SVD, è descritta nell'articolo *"Robust Model-Based Monocular Pose Initialization for Noncooperative Spacecraft Rendezvous"* di S. Sharma, J. Ventura e S. D'Amico **[31]**. L'articolo contiene un'esposizione teorica dei sottosistemi del metodo SVD e alcuni risultati preliminari che giustificano le scelte in termini di composizione dell'algoritmo. Questa implementazione dell'inizializzazione della posa comprende, in particolare, alcuni miglioramenti rispetto allo stato dell'arte consistenti in: una

tecnica di eliminazione dei gradienti deboli è posta in atto per determinare una regione d'interesse attorno al target all'interno dell'immagine, la definizione di parametri scalabili per la procedura di edge detection, un processo di sintesi delle features lineari che le classifica in gruppi concettuali sulla base di condizioni geometriche ed infine una consistente riduzione delle possibilità di matching basata su una combinazione più intelligente dei gruppi geometrici. I risultati presentati sono stati prodotti testando l'algoritmo su un ridotto insieme di immagini reali e artificiali del minisatellite Tango, parte della missione PRISMA.

Il corpo della tesi consiste nell'analisi dei risultati ottenuti dalla validazione dell'algoritmo SVD compiuta su due diversi dataset di immagini. La versione dell'algoritmo implementata nella tesi presenta alcune differenze rispetto a quella proposta nell'articolo originale, in particolare, in termini di risoluzione dell'equazione prospettiva dove al metodo EPnP viene sostituito il pose solver P3P+RANSAC per una migliore soluzione di posa. Il primo dataset, SPEED, è composto da 12000 immagini artificialmente generate del satellite Tango. Il dataset presenta una distribuzione normale di orientamenti a posizioni relative, sfondi in cui è presente la superficie terrestre e diverse condizioni di illuminazione, dunque, consente di definire una base statistica per valutare l'efficacia del metodo SVD. Nel terzo capitolo i risultati del test sono riassunti in alcuni grafici: il numero di stime di posa corrette e la loro accuratezza è relazionato ad alcuni parametri fondamentali come la distanza relativa, il numero di lati correttamente individuati e le combinazioni di features.

Nel quarto capitolo i risultati di un secondo set di immagini sono presentati. Il capitolo è diviso in una prima parte che descrive la creazione del dataset, la composizione del setup sperimentale all'interno della facility SPARTANS per la verifica della dinamica relativa tra minisatelliti e l'adattamento dell'algoritmo SVD al nuovo modello. Il setup consiste in un modello semplificato di un 2U-CubeSat, il quale ruota attorno ad un asse principale grazie ad uno stadio rotativo, ed una camera posta ad una distanza fissa. Come descritto nella seconda parte del capitolo, i risultati confermano le considerazioni precedentemente formulate sulle limitazioni del metodo SVD per la misura di posa.

In conclusione, la tesi esamina a fondo le peculiarità del suddetto algoritmo per evidenziarne i limiti intrinseci e i principali vantaggi rispetto ad altre tecniche basate su sistemi di visione monoculare. Il metodo SVD, infatti, si dimostra corretto nelle sue assunzioni, in quanto può ridurre notevolmente il carico computazionale della determinazione di posa con una precisione metrica centimetrica e precisione angolare nell'ordine del grado, però, a costo di una riduzione della percentuale di successo. In aggiunta, l'efficacia dell'algoritmo è fortemente dipendente dalla specifica

modellazione geometrica del target e dalla presenza di interferenze di fondo nell'immagine.

Introduction

The increasing need for autonomous rendezvous and proximity operations missions has driven the innovation in the field of relative pose measurements. Multiple types of current and future missions rely fundamentally on the efficiency of a robust guidance and navigation system (GNC), among them some key tasks like active debris removal (ADR), formation flying (FF), on-orbit servicing (OOS) require fast, reliable and adaptable measurement of relative range, line of sight and pose, i.e., orientation and position. The scenarios in which these missions take place can be various and the complexity of acquiring relative measurements can greatly increase on the basis of the cooperativeness and the passiveness of the target. Consequently, state-of-the-art navigation sensors and techniques can differ consistently in relation to the status of the target.

In the absence of a direct communication link between vehicles, a satellite is deemed non-cooperative and, besides LIDAR sensors, the most promising technology consists in imaging sensors such as monocular and stereo cameras. Given the continuous rise of computer vision techniques, the choice of camera pose estimation seems optimal in terms of computational cost, power and mass budget, robustness and hardware complexity. The implementation of vision-based algorithm is still secondary to the use of predefined visual markers which provide constant geometries and easily identifiable features to the detection algorithms embedded in the pose estimator architecture. In the first section of the thesis a brief overview of sensors and technique currently employed to perform relative pose measurements is presented

In regard to passive non-cooperative spacecrafts, in Chapter 2 an improved technique for pose initialization based on monocular systems is suggested. The so called SVD method comprises of an image processing section, a feature synthesis subsystem and pose solution and refinement. The method is described in S. Sharma, J. Ventura e S. D'Amico, "Robust Model-Based Monocular Pose Initialization for Noncooperative Spacecraft Rendezvous" [31] and some preliminary results are summarized in the paper. The main characteristics of the SVD technique are (1) the possibility to define a region of interest and provide a coarse estimation of range and line of sight, (2) an adaptable definition of the edge detector parameters, (3) a perceptual classification of

the detected segments in high-level groups and (4) a reduction of the search space based on the combination of complex and simple elements.

The main purpose of this thesis consists in elaborating further considerations on the effectiveness and the limitations of the SVD pose initialization. The central part of the thesis work comprises of the results of two validation tests: the SVD algorithm is implemented in the MATLAB software and tested on two different datasets. The first dataset expands the data of [31] with 12000 synthetic images of PRISMA mission's Tango spacecraft. The variety of attitude and translation values, together with the presence of Earth's images in the background, make up an optimal dataset for pose initialization testing. The validation results are analysed so to determine the elements that affect the success rate of the SVD method.

Furthermore, a CubeSat model is employed to create a second dataset of real images in more limited conditions of illumination and relative pose. The image set and the respective true pose values have been created at the SPARTANS testbed for relative satellite. The intent of the CubeSat experimental setup is to provide more statistical ground to the outcomes of the first validation test and to understand the degree to which the pose initialization technique can be improved by reducing the variance of some external factors.

Chapter 1

The problem of relative pose measurements

In this first chapter it's presented a survey of the state-of-the-art techniques and sensor systems employed to perform proximity operations. The discussion begins by addressing the scientific and technologic motivations that require the development of pose estimation architectures and it continues with a theorical description of the problem. The successive paragraphs provide a brief, complete overview of the most used navigation sensors and their peculiarities. Lastly, a categorization of the possible scenarios of cooperativeness for proximity operations is provided.

1.1 Motivation for relative pose estimation

Since the early stages of space exploration, the concept of rendezvous and proximity operations (RPO) has been studied and discussed by space agencies and private companies in an attempt to satisfy the continuously growing need for more adaptable, ambitious and complex space systems. From the Apollo program for Moon exploration to the servicing flights of the Space Shuttle to the assembly of the International Space Station (ISS) a number of successful missions have been enabled by the execution of relative approaching and cooperative procedures **[1]**. As it obviously appears, all these missions share a common trait: the operations were implemented in the context of manned and cooperative spacecrafts and, for this reason, the completion of the docking and servicing operations relied heavily on the astronaut's ability, along with the fine tuning of the guidance, navigation and control (GNC) system and continuous ground support.

Nevertheless, the increasing challenges of space exploration and especially the urgent need for efficient debris removal techniques has brought to attention the importance

of developing and testing autonomous and semi-autonomous rendezvous operations for non-cooperative spacecrafts. In almost any of the cross-cutting technologies and experimental concepts proposed for the present and next decade by manufacturers, satellite operators and space agencies, the execution of autonomous robotic proximity operations plays a pivotal role. The proliferation of satellites could unlock unthinkable capabilities in the commercial and civil field and, due to the increasing number of numerous small satellites constellations, the issues of space pollution, debris removal, non-functional spacecraft's deorbiting and safety rules for docking and formation flying will require the space sector to enact a technical revolution by promoting a new generation of completely autonomous unmanned systems.

As indicated by the National Aeronautics and Space Administration (NASA) Office of the Chief Technologist in 2011 the advancement of fully autonomous GNC systems is a fundamental requirement for the 2011-2021 decade and beyond **[2]**. The expected innovations must follow the following guidelines:

- *Improved Target State Estimation* : robust and accurate relative pose measurements for cooperative and non-cooperative spacecrafts.
- *Optimality* : optimal trajectory planning to minimize energy and fuel consumption.
- *Robustness* : reliability over orbital perturbations and sensor noise.
- *Real-time and on-board implementation* : The algorithm must be sufficiently fast and computationally inexpensive.
- *Manoeuvring Precision*

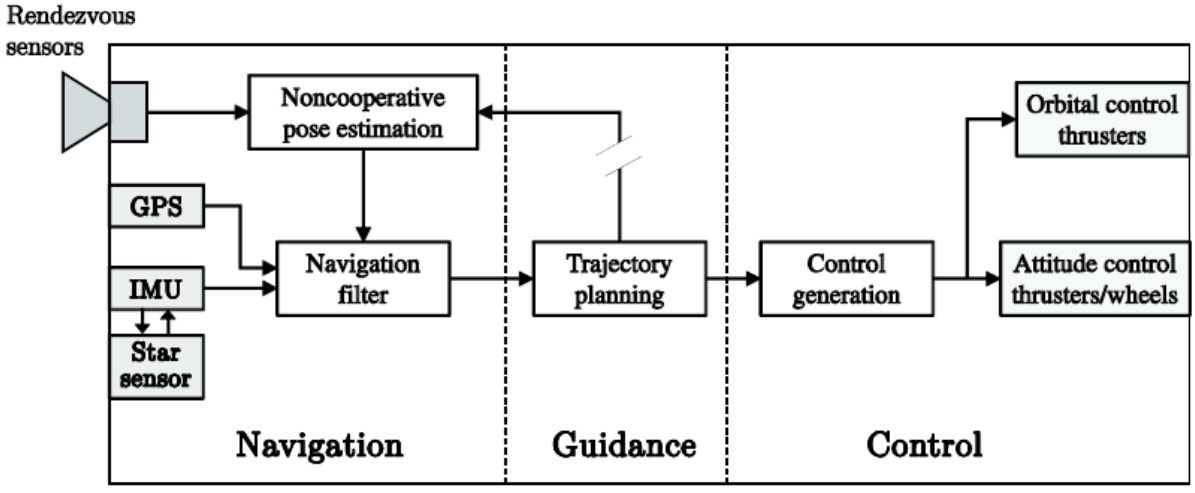


Figure 1.1 Schematic overview of a GNC system for autonomous noncooperative proximity operations [3].

A conceptual architecture for a modern GNC system is shown in **Fig. 1.1**. The input measurements acquired by absolute (star trackers, horizon sensors, GPS receivers, etc…) and inertial (IMU) attitude sensors are combined with the relative attitude and position data as elaborated by a pose estimation algorithm. The algorithm relies on the rendezvous sensor suite which can be composed of a variety of different instruments, such as LIDAR sensors, radar sensors or stereo and monocular cameras, whose peculiarities will be discussed further in the thesis. Consequentially, the data is fed to an Extended Kalman Filter which generates navigation information and filters measurement noise. The guidance segment is dedicated to the task of planning the optimal trajectory to the target while the control segment is divided into a parametrization software for attitude control and the on-board actuators, such as thrusters and reaction wheels **[3]**.

Although most components of the logical architecture presented above have already been developed separately for space applications, there is still a technologic gap to be closed in order to achieve fully autonomous guidance system and particularly in reference non-cooperative pose estimation, which is discussed in the thesis.

1.2 Problem statement

Relative pose estimation is the problem of computing position and attitude of a target space object in relation to an active spacecraft that acts as chaser. The issue results in identifying unambiguously the orientation of a Target Reference Frame (TRF) with respect to a Sensor Reference Frame (SRF) by a set of six orbital parameters representing the translational and rotational degrees of freedom. During far and mid-range operativity, however, given the large distances, only Line of Sight (LOS) and relative range measurements are required to ensure mission requirements. A series of different maneuvers can be performed by a chaser: coordinated orbiting, translations along flight axis, monitoring, station-keeping, close approaching, docking and, generally, operations that extend from tens of kilometers to contact. Pose determination is a complex task and represents, in any scenario of proximity operativity, a key feature to achieve autonomous navigation accurately and safely.

The quality of relative pose estimation depends fundamentally on the type of technological solution employed to provide measurements, the cooperativeness of the target and the mission's scenario. In particular, in the next paragraphs the issue of cooperativeness and its influence over mission architecture are discussed. Furthermore, a brief survey of instruments and techniques for pose estimation is presented.

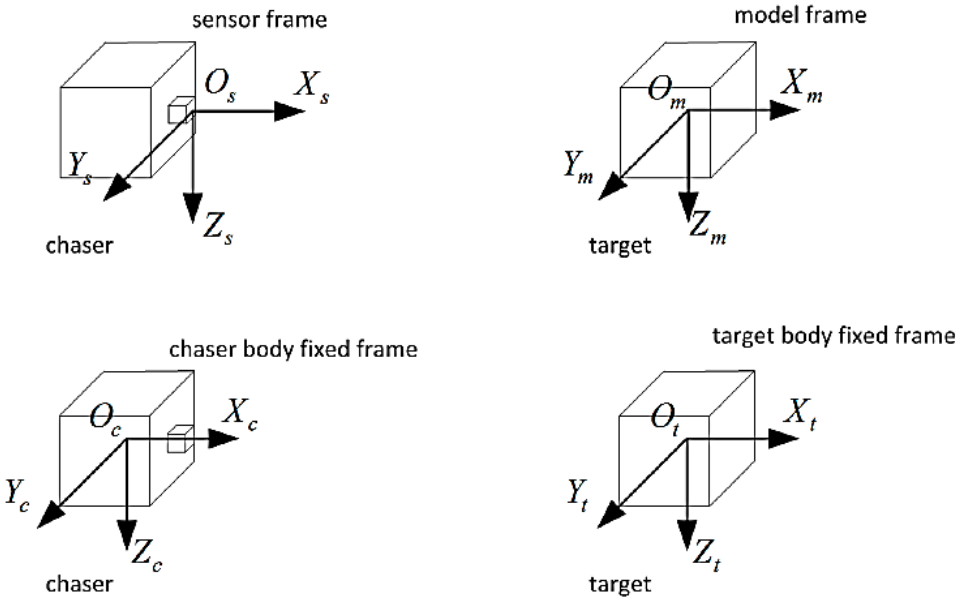


Figure 1.2 Typical reference frames.

For the problem of pose measurements four reference frames are of interest: the chaser body-fixed frame, the sensor frame, the target model frame, the target body-fixed frame, as shown in **Fig. 1.2**.

Chaser and target body-fixed frame usually lie in the center of mass of both satellites and their orientation is formally defined by the orbit and attitude control systems. The origin of the sensor reference frame O_S-$X_S Y_S Z_S$ depends on the sensor architecture, for instance in the case of imaging detectors the camera frame lies in the center of perspective projection. Lastly, the target model frame is fixed in respect to the body frame and often coincides with it, it is typically determined in relation to fiducial markers locations so to allow simpler perspective calculation. Assuming that the transformation matrix from the sensor frame to the chaser body frame is known by design and similarly the transformation matrix from the target's model frame to body frame, the relative attitude and location data that are required for navigation control are composed by the rotation matrix and translation vector which describe the transformation from the sensor frame to the target model frame.

A transformation between reference frames defines unequivocally the pose and it can be expressed by a roto-translational matrix. The translation vector T, as in **Eq. (1.1)**, represents the relative position between O_S and O_M and the rotation matrix R is determined, as in **Eq. (1.2)**, by a 312 sequence of elemental rotations about three linearly independent axis by three angles (ϕ, θ, ψ) according to the Euler angles definition. Considered a point in the model frame $P_M = (x_M, y_M, z_M)$ and the

corresponding point in the sensor frame $P_S = (x_S, y_S, z_S)$, **Eq. (1.3)** and **Eq. (1.4)** are satisfied and equivalent.

$$T = [\Delta x, \Delta y, \Delta z] \tag{1.1}$$

$$R = R_Y(\theta)R_X(\varphi)R_Z(\psi)$$
$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

(1.2)

$$P_S = R \times P_M + T \tag{1.3}$$

$$[P_S\ 1] = [P_M\ 1] \times \begin{pmatrix} R & 0 \\ T & 1 \end{pmatrix} \tag{1.4}$$

1.3 The issue of cooperativeness

The relative pose estimation problem is a challenging priority for the future of proximity operations between spacecrafts and a crucial element to its feasibility is the cooperativeness of the target. A target is defined non-cooperative if doesn't provide any direct information about its motion and can't, actively or passively, facilitate rendezvous manoeuvring. Obviously, multiple levels of cooperation can be distinguished on the basis of prior state knowledge, motion stability, possibility of remote control, presence of fiducial markers, ability to implement inter-vehicle communications and more.

As shown in **Fig. 1.3**, which presents a brief summary of on-orbit servicing missions, there is already a successful history of cooperative docking procedures, especially in the case of International Space Station's refurbishment missions. After years of continuous development and progress, at the moment the ISS performs docking and berthing maneuvers roughly every two months with manned or unmanned vehicles that operates in partial autonomy. The fact that several distinct spacecrafts have accomplished the goal of approaching and connecting with the Station stands as proof that the current state of GNC system for cooperative system is sufficiently mature.

Nevertheless, the ISS is an exceptional case since it receives full operational support from the ground segment, each docking port is equipped with LIDAR sensors, thermal cameras and HD cameras to provide a visual feed to astronauts aboard and, also, the ISS is stabilized with three-axis attitude control.

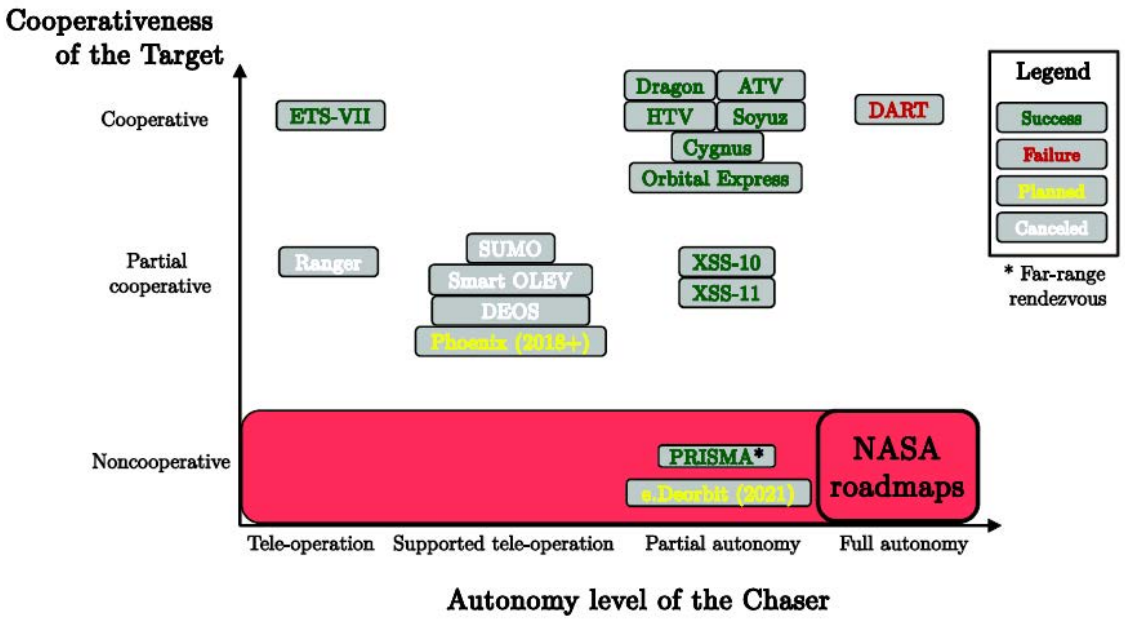


Figure 1.3 Overview of some space missions that execute proximity operation and docking procedures **[3]**. The diagram is updated to 2016.

Other experimental missions, mostly successful, have been launched in the last decades and progressively more advanced and autonomous guidance systems have been tested with less cooperative targets, demonstrating how the absence of cooperativeness exponentially increases the complexity of servicing operations at close proximity and requires intelligent and precise sensors.

A first attempt at proximity procedures was the DART (Demonstration for Autonomous Rendezvous Technology) mission which goal was to perform independently an initial approach maneuver, station keeping at 5 m from the target and a series of collision avoidance simulations. The target of DART operations is a previously released communications satellite (MUBLCOM) orbiting at 775 km of altitude, purposefully outfitted for a possible cooperative relative pose estimation. DART's sensor suite is composed by a laser-based optical system that, when coupled with dedicated reflectors on the target's structure, allow for a relatively accurate estimation of attitude and position. Furthermore, it receives a continuous GPS information stream from the MUBLCOM via an intersatellite UHF communication link **[5]**. Unfortunately, the mission failed to achieve its goals due to guidance errors

which led to incorrect propellant management and, therefore, to a soft collision between the two vehicles.

After DART mission failures the US AFRL (Air Force Research Laboratory) launched XSS-10 in 2003 and shortly after XSS-11, two microsatellites that managed to perform inspection tasks and prove automated co-orbiting capabilities. In contrast to DART, both spacecrafts were maneuverable from ground control sites and were equipped with imaging sensors. While XSS-10 was focused on demonstrating navigation around a single Delta-2 second stage burner on a pre-planned course **[6]**, XSS-11 successfully orbited multiple satellites during a yearlong life span and adopted an advanced autonomous event planner for onboard iterative trajectory planning, thus, greatly improving the results of previous missions. Since the XSS-11 spacecraft had to perform inspections of non-cooperative RSOs (Resident Space Object) a main asset of its sensor suite was the Rendezvous Laser Vision (RLV) instrument which consisted in an integrated active LIDAR system capable of detecting and tracking object **[7]**.

In an attempt to further explore the feasibility of a wide range of newly developed relative attitude sensors for formation flying, the PRISMA (Prototype Research Instruments and Space Mission technology Advancement) spacecraft was launched in 2010 **[8]**. The mission comprises two vehicles: MAIN (also known as Mango), a highly maneuverable 3-axis microsatellite with reaction wheel-based attitude stabilization and 3-axis ΔV capability with six 1 N hydrazine thrusters, and TARGET (also known as Tango), a smaller spacecraft with coarse magnetic attitude control but without ΔV capability. Given TARGET's ability to simulate different degrees of cooperativeness, mission requirements were focused on MAIN performing several station keeping, approaching and rendezvous tasks in autonomy while correctly assessing Tango's pose with multiple instruments. Along with complex close-loop formation flying algorithms, PRISMA's Mango relied on three fundamental tools for relative motion estimation **[9]**:

- *Differential GPS* for precise orbit determination and long-range operations in cooperative state.
- The *FFRF (Formation Flying Radio Frequency)* metrology subsystem which offers sufficiently accurate measurements of position, velocity and line of sight, besides intersatellite communication link.
- A *Vision Based Sensor (VBS)* consisting of multiple star cameras able to identify non-stellar objects and determine both pose and range in non-cooperative states.

The characteristics of these techniques will be discussed in the next paragraph but the value of PRISMA's experimental activities remains, still now, essential to the advancement of a variety of future missions where rendezvous and formation flying are a necessary prerequisite. In fact, the first part of this thesis originates from the work on vision-based pose estimation that follows the results of PRISMA and hinges on the data produced in conjunction with the mission.

1.4 State-of-the-art pose estimation sensors

An accurate and trustworthy solution to the pose estimation problem represents a key element of any navigation system and it's crucial for the advancement of GNC systems. As mentioned above, the target cooperativeness has an impact on the selection of the specific sensor architecture as well as on the pose solver algorithms.

The rendezvous sensor suite for close proximity operations can be differently composed depending on the mission's complexity and spacecraft's requirements in terms of mass and power budget. Throughout the history of OOS missions various technological solutions have been proven to be effective and sufficiently accurate and they are described in the following paragraphs.

1.4.1 Radio Frequency (RF) antennas

RF metrology has been exploited in early rendezvous missions by space agencies and provides range, LOS angles and indirectly relative attitude parameters, while operating from a distance of several kilometres down to contact **[4]**. A fundamental limitation to this system was represented by the power and mass requirements, yet more recently the FFRF system, tested on PRISMA, proved the feasibility of a low-budget metrology set-up with a total of 3 Rx/Tx antennas for TARGET and a full antenna triplet for MAIN. PRISMA RF subsystem is able to provide relative positioning of 2 to 4 satellites in formation flying and consists of one terminal and up to 4 antennas for each vehicle in a constellation **[10]**.

1.4.2 Global Navigation Satellite System (GNSS) receivers

Even though this solution is not suitable for deep-space missions, the presence on board of GPS receivers and antennas when paired with an intersatellite communication link allows for a precise calculation of relative position through the Differential GPS or Carrier-phase Differential GPS techniques. The network of GNSS satellites offers continuity of signal and global coverage for every spacecraft in Low Earth Orbit (LEO) nonetheless, in case of operations within a relative range

of 10 m, measurements are greatly affected by the phenomenon of multipath interferences and partial occlusion **[4]**.

1.4.3 Light Detection and Ranging (LIDAR)

Every instrument that measures distances by illuminating a target with an infrared laser and analysing the backscattered radiation falls under the definition of LIDAR. For spaceborne applications, the most common choices for sensors are scanning Lidar systems and detecting arrays systems. While the first ones rely on high-speed and high-precision optical mechanisms to convey a narrow laser beam and detect the reflected light in predetermined pattern, the second ones illuminate the scene with a broad laser beam and use a detector array to absorb the backscattered echoes in the pixel direction **[11]**. LIDARs produce sets of 3D data (point clouds) which can be elaborated for target tracking and pose estimation. Regardless of specific weaknesses such as having conspicuous power consumption, requiring large amounts of memory and computational power on board and having a typically poor spatial resolution, LIDARs are robust to operations in any visibility condition, they allow for an easy discrimination of the target from the background and, lastly, they provide both 3D position and intensity maps. Furthermore, LIDARs can be divided into pulsed and continuous-wave (CW) sensors on the basis of light source's characteristics. CW detectors usually operate in close range operations (below 15 m) since, by exploiting the phase shift between emitted and absorbed radiations, the measurements suffer from the integer ambiguity problem and the accuracy decreases as the inverse of the square root of the distance **[11]**. However, when in the classical triangulated configuration, this system can ensure millimeter or even sub-millimeter accuracy precision. On the other hand, pulsed LIDARs are Time-Of-Flight (TOF) instruments that compute range by means of measuring time delay between radiation emission and absorption, given the constant of light speed. TOF sensors represent the optimal choice for proximity operations over a large interval of distances, namely, from 1 m to several kilometres with an accuracy that varies only as a function of the highest achievable time resolution.

A mentionable example of pose acquisition based on LIDAR is the LIRIS-2 sensor mounted on the servicing vehicle ATV-5 to the International Space Station. The system records a high-resolution 3D measurement of the ISS during the approach. The instrument has been tested in the close range from 1 km down to contact and thanks to the presence of dedicated reflectors, the mission has been successfully concluded by a correct docking procedure [.

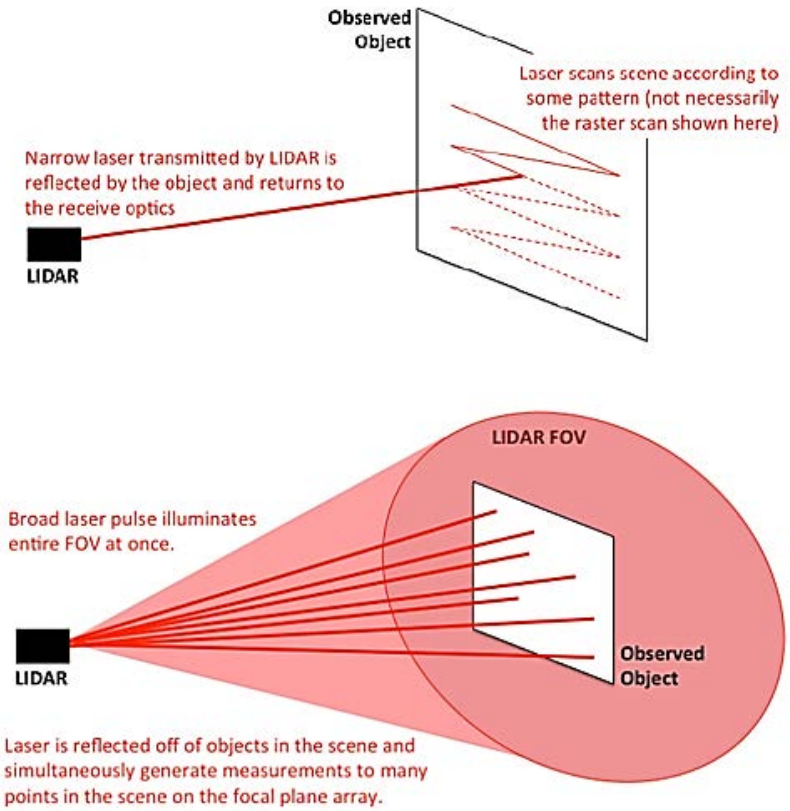


Figure 1.4 Notional depictions of LIDARs in scanning and detector array configuration. **[11]**

1.4.4 Monocular and stereo cameras

As space vehicles are rapidly evolving towards more compact, efficient and intelligent designs the role of imaging sensors has become crucial for pose estimation because of their lower hardware complexity, cost and power consumption. Typically, infrared cameras are primarily employed to perform LOS measurement at far range since the data is too poorly textured for pose determination purposes, whereas sensors that work in the visible range (between 0.37 μm and 0.75 μm) offer more reliable image quality at various relative distances. Moreover, images are fundamental sources of information for supervised applications, such as ground control situational awareness and human-in-the-loop operations. Monocular cameras have been proven to be suitable for wider operational ranges, especially when coupled with fiducial markers and reflecting elements on the target or LEDs emitting visible light installed on the chaser. In regard to stereo cameras the feasibility of 3D reconstruction and depth resolution for cooperative and non-cooperative spacecrafts has been tested in formation flying demonstration missions but accuracy and range are strictly related to the baseline which must be maximized with respect to chaser's dimensions. The

main drawback of imaging sensors is their unpredictability when exposed to largely variable illumination conditions or to direct light from other celestial bodies, not an uncommon scenario in space. Compared to LIDARs, monocular and binocular systems provide highly processable and easily readable data, they are less prone to mechanical failures and convenient in terms of weight, power consumption and required processing power. Even though imagers can singularly lack flexibility in various operational contexts due to constant optical proprieties (FOV, focal length, pixel density, …), usually rendezvous sensor suite rely on multiple collaborative cameras, each one specifically designed to be activated during distinct phases of the approach maneuver.

1.5 State-of-the-art pose estimation techniques

As discussed in previous paragraphs the choice of an adequate technological solution to this problem is profoundly influenced by three fundamental factors:

1. Sensor suite composition
2. Mission requirements
3. Target cooperativeness

A wide variety of hardware architectures and navigation algorithms have been developed for pose estimation throughout the history of space exploration, yet they fall into four separate categories that summarize all the possible scenarios which require autonomous operativity **[4]**. Multiple state-of-the art techniques, divided by mission scenario, are presented in the next paragraphs.

1.5.1 Actively cooperative target

The definition of actively cooperative target refers to a known vehicle with the ability to communicate its orbital parameters and its attitude to other spacecrafts or to a common ground segment. In addition to these properties, the target spacecraft is usually supposed to be stabilized via autonomous dynamic control or remote ground support, thus, requiring accurate and redundant attitude determination sensors and an efficient attitude control system. Currently, the vast majority of operating space vehicles is equipped with star trackers that guarantee attitude estimation with a precision inferior to 10 arcsec and IMUs that, in different configurations, can provide accurate inertial navigation at drift rates as low as 0.001°/h, moreover, thrusters and reaction wheels provide trustworthy and flexible technologies for motion stabilization. Therefore, once a communication link is established, every spacecraft can rely on on-board attitude determination to enable relative pose estimation and

proximity operations. The scenario of actively cooperative satellites concerns typically Formation Flying (FF) missions: whether in trailing configuration or in clusters, orbital dissipation tends to alter the required relative position and velocity between spacecraft and must be opposed through active orbit control, which has progressed towards a more autonomous approach, as space constellation become more numerous and require tighter maneuvering.

Relative pose estimation with *RF communication* is historically the first elementary method to adequately perform rendezvous operations **[12]**. For instance, the approaching and docking of Progress and Soyuz vehicles with Russian space station Mir and, later, with the ISS was completed in various missions by establishing radio communication between the two manned vehicles and exchanging attitude and position information. A fundamental evolution in RF metrology has been tested with the FFIORD (Formation Flying In-Orbit Ranging Demonstration) on board of PRISMA's main spacecraft. As reported by J. Harr, M. Delpech et al**. [10]** the experiment aimed at assessing measurement performances while conducting various automated activities like standby in safe relative orbit, translation maneuvers in plane and cross track, station keeping at multiple distances and collision avoidance. Measurements of LOS and range are obtained directly with an approach inspired by the GPS system, i.e., each vehicle transmits and receives a GPS C/A navigation signal modulated on two S-band carrier frequencies, once transmitted, RF signals firstly allow the chaser to formulate a coarse evaluation by means of pseudo-code ranging and secondly carrier-phase difference estimations increase accuracy to less than 10 cm. The intersatellite link also provides the relative clock drift of the platforms and attitude data as calculated by on-board sensors. Results from ~75 days of testing procedures show that the RF subsystem can achieve, in the range between 3 km and 30 km, a precision of 2 cm for relative distance measurement and 1° for LOS measurement, yet multipath errors produced, respectively, mean errors of 90 cm and 20° when target elevation on the chaser orbit plane is greater than 35° **[15]**.

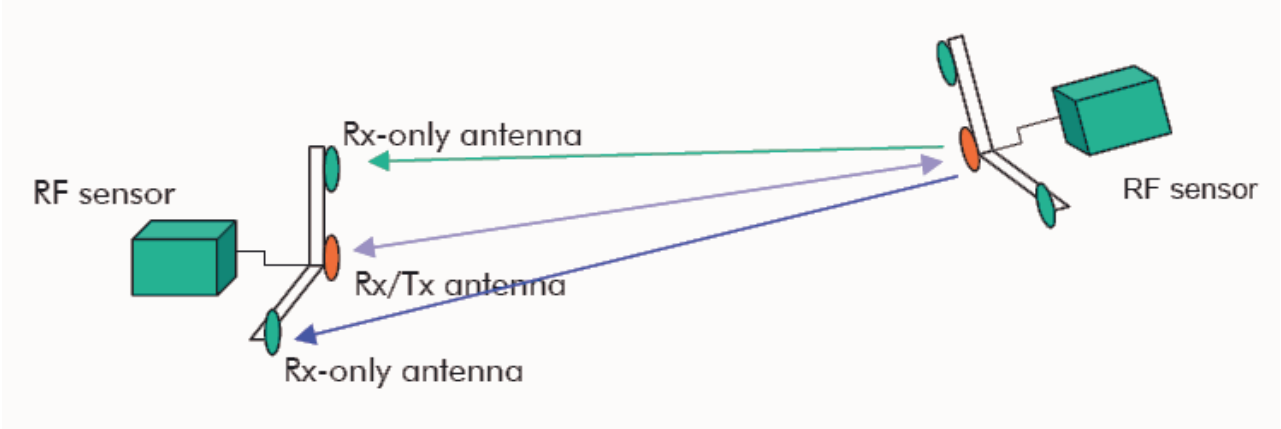


Figure 1.5 Two satellite FFRF antennas configuration. **[15]**

PRISMA's FFIORD experiment proved that RF pose estimation serves as a sufficiently precise and robust technology especially for future formation flying missions, nevertheless it's affected by fundamental disadvantages such as mass and power consumption, multipath errors, ionosphere interference and relying on board sensors for attitude estimation.

A robust and effective solution for the problem of relative position and pose estimation of active cooperative spacecrafts is represented by GPS based measurements. The Global Navigation Satellite System (GNSS) is a satellite navigation system that relies on multiple constellations of spacecrafts (GPS, GLONASS, Galileo, BeiDou, …) that provide position data with a continuous global coverage. GPS receivers are lightweight, reliable and particularly cost-effective instrumentation for LEO missions, in fact, GPS signal can be manipulated through various techniques to obtain different levels of accuracy for relative position. Differential GPS measurements provides position information with a precision inferior to 1 m while Carrier-Phase Differential GPS can further improve on accuracy reaching centimetre-level precision **[13]**. For instance, PRISMA's MAIN and TARGET spacecrafts are equipped with a Phoenix GPS receiver qualified by DLR and flight-proven in previous missions (GRACE, CHAMP, Proba-2, …). With a mass of 70 g and a power consumption of 0,85 W the receiver is suited for small satellites and it offers single-frequency code tracking for the dynamic modelling of relative orbits and carrier phase measurements for more precise intersatellite distances **[14]**. GPS signal doesn't provide attitude estimation therefore any pose determination must rely on on-board attitude sensors and communication with the ground segment for remote control or space-to-space communication links for autonomous flying. Despite its many advantages GNSS based relative positioning is affected by multipath errors, signal occlusion, ionosphere interferences and radiation-induced hardware failures, for these reasons it can only be suitable for mid/far range proximity operations and LEO missions.

1.5.2 Passively cooperative target

The most common scenario of On Orbit Servicing missions involves passively cooperative target vehicles, i.e., space objects equipped with artificial markers that offer different types of visual aid for electro-optical rendezvous sensors. Vision-based pose estimation appears to be the most effective in relation to target passiveness since it doesn't require intersatellite data exchange and, as described previously, it has a minor impact on power and mass budget. Both in the visible and infrared range optical sensors rely on pre-determined markers to be visible and to produce a constant, robust response to imaging detection. Given the a priori knowledge of the position of a limited number of features in the Target Reference Frame and either the

2D coordinates on the sensor's image plane or the 3D coordinates in the sensor reference frame of the same features, it's possible to retrieve the relative orientation between the two vehicles. Fiducial markers must have invariant and easily identifiable geometric correlations which can be exploited to calculate coordinates and distances among features in the projective space.

Regarding monocular cameras different techniques have been proposed based on various types of fiducial markers. The use of optical sensors to obtain line-of-sight vectors has been explored by Junkins et al. **[16]**, in Philip and Ananthasayanam **[17]** and in numerous other related works that explore the feasibility of computer vision in space applications. A first example of vision-based pose estimation flew on board of the Engineering Test Satellite (ETS-VII) autonomous docking mission where the Proximity Operation Sensor provided reliable navigation data in the range between 3 m and contact, based on the acquisitions by a single camera of seven non-coplanar, round shaped passive markers on the surface of the target spacecraft. Also, Ho and McClamroch **[18]** propose a rhombus shaped marker for autonomous soft docking. Another technological solution for fiducial markers is represented by Corner Cube Reflectors (CCR) which can be precisely built to reflect light at specific wavelengths, thus, enabling background-subtraction method to detect markers. In fact, the Advanced Video Guidance Sensor on board of DARPA's Orbital Express mission produced light emissions at 808 nm and 845 nm, yet its target's CCRs only filter the former one. By subtracting the resulting intensity images the background radiation, which appears in both, is removed and only markers reflections remain. Once identified, valid spots are selected as inputs for an Inverse Perspective algorithm **[20]** in order to estimate the relative pose. Mission reports **[21]** highlight the excellent results of this technique at short range: the AVGS subsystem provided an accuracy of 0.75 mm (one order of magnitude better than the requirement) while operating in mated configuration at a true range of 1.22 m.

Furthermore, the aforementioned PRISMA experimental spacecraft was equipped with the VBS optical subsystem for vision-based pose estimation which is composed of four imaging sensors operating, respectively, two star trackers (main and redundant), a far range and a short range monocular sensor for relative measurements. The system relies on several switchable LEDs mounted on the target spacecraft as fiducial markers and exploits their unique geometrical pattern to perform 2D-to-3D point matching **[8]**. The VBS Far Range camera demonstrated to be able to detect and track multiple non-stellar objects at distances up to 500 km and provide measurements of LOS and relative motion, while the Close Range camera provided sufficient resolution to extract the coordinates of luminous elements and detect predetermined patterns in order to assess relative attitude and position with an

accuracy of 1 cm and 1° for at 10 m range **[15]**. Although the optical equipment includes a pass band filter centered on the LED wavelength, the marker detection technique is fundamentally affected by illuminating conditions.

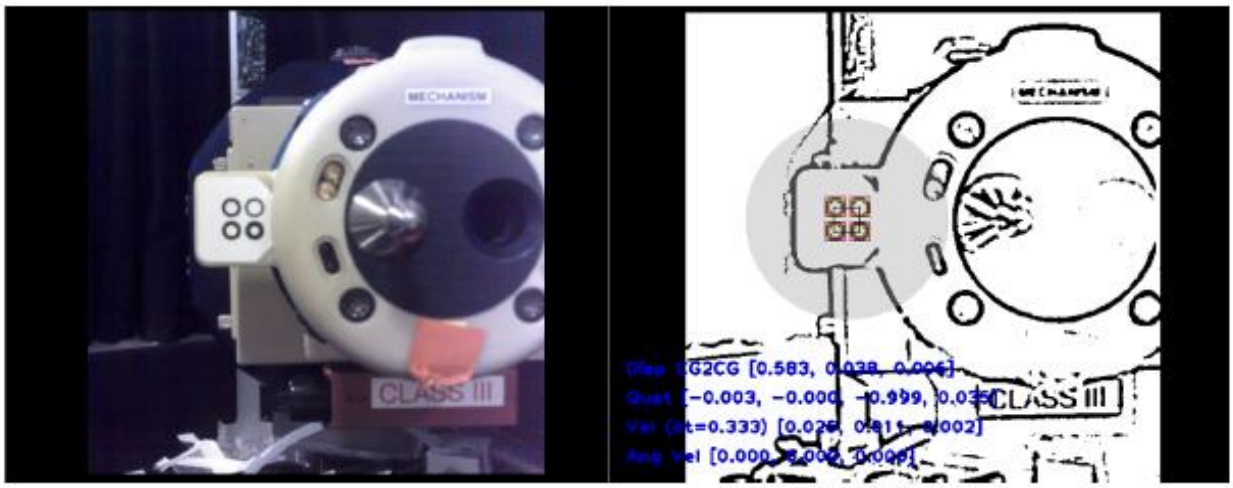


Figure 1.6 SPHERES fiducial markers (left: raw image, right: processed image and markers detection). **[23]**

Sklair and Gatrell in **[22]** identify concentric contrasting circles as the optimal geometry for passive fiducial markers in relation to space lighting, ease of manufacturing and robustness against image distortion. In fact, a set of four stacks of two concentring black and white circles is employed to perform close proximity operations among the experimental microsatellites of the SPHERES program aboard the ISS. Within the SPHERES project, developed by MIT Space Systems Lab, Tweddle **[23]** proposed an algorithm architecture for pose estimation which comprises of image processing for fiducial detection, solving the exterior orientation problem and filtering the measurements through a Multiplicative Extended Kalman Filter (MEKF). In particular the feature detection process revolves around four phases: masking the image around a previously tracked location, thresholding the image to segment it into black and white regions, implement a blob detection algorithm and measuring their geometric properties (perimeter to area ratios, eccentricities, areas, relative distances …) to remove the wrong ones and, finally, calculating the centroid's 2D coordinates to feed to the pose solving algorithm. The technique ensures an error of less than 0.5 cm for translations of circa 40 cm and a maximum error of 4° for rotations of 45° around the vertical axis **[24]**.

An alternative approach to pose estimation is provided by 3D measurement sensors such as LIDARs and stereocameras, although these techniques are mostly employed in the case of non-cooperative spacecrafts due to their operational flexibility. While

in **[25]** an algorithm architecture for stereo-vision autonomous navigation at close range is presented, yet a specific solution to detection and identification of fiducial markers is not addressed, there is a rich literature about Flash LIDAR application in proximity operations. A method to create uniquely distinguishable 3D point patterns for fiducial LIDAR measurements can be found in **[26]** and a graph-matching algorithm for reflector identification in **[27]**. In the majority of cases data analysis comprises of thresholding and filtering operations to be applied to the LIDAR intensity image based on morphology. Thereafter, pose estimation relies on a linear formulation of the 3D point pattern aligning problem which has been proved to obtain accuracy around 1° and 1 cm at distances from 2 m to 20 m **[4]**.

1.5.3 Non-cooperative known target

Spacecrafts involved in Active Debris Removal (ADR) missions typically belong to the category of non-cooperative known vehicles since they are defunct satellites or uncontrollable debris devoid of fiducial markers or active communication systems, whose structural and optical properties are known and can be at least partially reproduced. In this scenario the pose determination process can be divided into an initialization phase and a tracking one. Lacking any a-priori knowledge of target attitude, the first acquisitions allow estimating the pose parameters by matching measurements data with the respective information derived from a target model stored on board. The consecutive step, instead, consists of pose tracking algorithms that update and refine the stream of relative measurements on the basis of previous pose estimates with the goal of providing a more robust and accurate description of the target's motion.

Generally during the acquisition phase, the feature matching techniques rely on approaches, commonly used in computer vision, to extract and characterize geometric elements, like corners, lines, shapes and curves, from 2D or 3D point data. The SoftPOSIT algorithm **[28]**, based on the Pose from Orthography and Scaling (POS) process, is employed to search for image-to-model correspondences, firstly it matches feature with a soft-assign technique, then it defines an initial coarse pose estimation by means of a scaled orthographic projection and lastly it computes a refined solution iteratively. SoftPOSIT can also be applied to lines, producing more consistent solutions due to their inherent stability to lightning conditions and invariancy to clutter and noise. Feature matching can be performed also with the RANSAC (Random Sample Consensus) method as described in **[29]**. Moreover, Template Matching is presented in **[30]** as a viable solution to identify and match image sections by exploiting segmentation and correlation functions based on the sum of absolute differences, the normalized cross correlation or the distance transform. Lastly, more recent architectures have been tested to apply perceptual feature grouping to the

problem and therefore reduce the computational cost of pose estimation while ensuring a more robust solution, indeed, the method proposed by Sharma, Ventura and D'Amico in **[31]** is the subject of this thesis and will be broadly discussed in the next chapters. Optimal monocular approaches can achieve attitude accuracy close to 1° and position error close to 1% of the true relative range **[4]**.

While 2D feature properties tend to be greatly affected by external factor and difficult to associate with a predetermined model dataset, 3D features rea more easily identifiable, especially in the case of LIDAR measurements. Given a 3D representation of the target, in addition to feature descriptor methods, point cloud techniques provide computational time saving and robustness to outliers, distortion and noise. Among several algorithms the Iterative Closest Point (ICP) registration method is typically used to determine the rigid transformation that aligns two datasets by optimizing a cost function, e.g., sum of squared differences between the coordinates of matched pair. Furthermore, ICP is easily customized to the specific application of interest by modifying some parameters and, in many cases, it has proven to determine a correct relative pose with centimetric and sub-degree accuracy in position and attitude, respectively.

1.5.4 Non-cooperative unknown target

Uncontrolled debris and asteroids, known as Resident Space Objects (RSO), belong to the category of non-cooperative unknown targets. Asteroid rendezvous missions and debris removal operations represent the most complex and risky scenarios for navigation purposes. Pose determination for requires a preliminary monitoring phase that involves building a model or a map of the target directly on board using the available measurements. In order to accomplish this task, solutions based on Simultaneous Localization and Mapping (SLAM), widely explored in robotic vision-based odometry, are applied. SLAM consists in the real-time process of establishing the position of a mobile robot while simultaneously building a map of the unknown environment. SLAM techniques have been studied in the last 30 years and have been employed successfully on board of Mars rovers to perform local trajectory verification and tested in the case of asteroid robot exploration **[58]**. Even though, some fundamental hypothesis must be verified, e.g., quasi-static scene, free-floating rigid objects and sensor synchronization, it is possible to adopt feature detection and optical flow computational techniques to estimate relative position. Augenstein and Rock **[32]** propose a monocular vision SLAM approach that provides pose tracking and shape reconstruction by means of Bayesian filtering and measurement inversion. The algorithm operates on the assumption of a constant rotation al velocity. Other implementations of the same architecture rely on both 3D sensors and cameras to produce combined measurements and, in addition to mapping and tracking the target,

can identify its inertia tensor **[33]**. Finally, the issue of proximity operations with non-cooperative unknown targets is still considered unsolved due to the lack of sufficiently fast and accurate implementations of SLAM, despite the numerous possible combinations of sensor suites, filtering processes and dynamic models that can be arranged, there isn't still a consistent and safe method to provide navigation data within such mission requirements.

Chapter 2

The SVD architecture for pose initialization

The chapter provides a detailed description of a model-based pose initialization technique for non-cooperative passive target vehicles. The so-called SVD architecture comprises of multiple subsystems for feature detection and synthesis, pose solution and refinement and the chapter points out the fundamental instructions to implement the algorithm. A summary regarding the preliminary validation results of the SVD method, contained in an original study paper, is outlined.

2.1 Architecture overview

One of the most promising and possibly advantageous technologies for on-orbit servicing and active debris removal with non-cooperative targets is monocular vision. As described in the previous chapter, monocular navigation guarantees fast pose determination under low mass and power requirements and can potentially provide reliable data for real-time generation of 6 DOF approach trajectory and control update if enough computational power is available. While proximity operations missions with unknown non-cooperative RSOs still require more precise relative measurements for target mapping and, therefore, rely necessarily on a combination of LIDAR and optical sensors, there are concrete possibilities for vision-based techniques to be robust and accurate enough to be employed singly with known non-cooperative targets. It's important to highlight that many algorithms to solve the pose problem are inherited from previous studies in computer vision terrestrial applications.

Vision-based monocular pose estimation architectures generally comprise of three fundamental blocks:

- *Image processing*
- *Pose solving*
- *Pose refinement*

The processes of prefiltering, feature detection and feature matching are part of the image processing phase and they ensure the extraction of significant 2D points and elements and provide a coarse estimation of target's range and line of sight. Due to the highly variable illumination condition, the high reflectivity of commonly used space materials, high contrast and low signal-to-noise ratio image processing is the main challenge of pose estimation. The first goal of the processing procedure consists in identifying and defining a Region of Interest (ROI) for the observed target, yet external factors like the presence of a reflective background (i.e. Earth surface) or the lack of sufficient light (i.e. Sun eclipse) can greatly affect the effectiveness of feature detection algorithm. Given the variety of conditions under which images are captured, the delimitation of a correct ROI allows to avoid quantitative biases in most feature detection techniques, since they can be manually tuned for better accuracy.

State-of-the-art techniques for feature extraction include corners detectors such as FAST (Features from Accelerated Segment Test) **[34]** and Harris **[35]**, extractors based on blobs detection like SURF and SIFT, binary robust invariant scalable key-points (BRISK) detector **[36]** or various edge detectors (Canny **[37]**, Sobel **[38]**, Prewitt **[39]**) paired with the Hough transform. If the pose estimation algorithm isn't correctly initialized, by means of Kalman filter state prediction or secondary relative measurements, the aforementioned techniques can't consistently discriminate by default the true features of the target from background interference and require a subsequent verification algorithm to evaluate the correctness of feature correspondences between image and model. Dhome et al. **[40]** proposed a closed-form solution by combining all the possible sets of three 2D image lines with 3D spacecraft model edges, yet more recently other authors have opted for RANSAC **[41]** and soft assign algorithm **[42]** to reduce the computational burden. Furthermore, as described in previous paragraphs, hierarchical view graphs and template matching represent more effective pose initialization architectures, however, they rely heavily on the presence of an exhaustive database of precomputed renderings of a target's model whose quality, in terms of illumination predictability and variety of conditions reproduced, can limit the extent of these approaches. Finally, more flexible solutions to the same problem have been proposed, i.e., a combination of background subtraction and Gaussian blob detection algorithm or maximal stable extremal

regions (MSER) detection **[43]** with non-maximum suppression (NMS) **[44]** which provide the possibility of identifying multiple spacecrafts in the image. In any case, the classical methods for image segmentation and target identification are computationally expensive and lack consistency in results.

Once the correspondences between extracted features and model features are assessed, the estimation of the position and orientation of a calibrated camera from n 3D-to-2D point correspondences is defined as the Perspective-n-Points problem. PnP algorithms require a minimum of 6 correspondences for a unique solution but can provide a set of 4 ambiguous solution for only 3 correspondences. A wide variety of perspective equation solvers have been proposed for the PnP problem with different limitations: among the non-iterative solvers the P3P algorithm described in **[45]** and the EPnP method described in **[46]** are the most commonly used in computer vision applications. The P3P approach is based on the smallest subset of control points (n = 3) that yields a finite number of computed poses and, therefore, it can be employed in combination with a RANSAC outlier detection for sets of easily identifiable features such as with fiducial markers. Since the computational complexity of perspective equation solvers grows exponentially with the number n of matched features, the EPnP algorithm, which provides a better accuracy with linear complexity $O(n)$, has optimal performances for larger values of n ($n > 4$) reducing sensitivity to noise, without sacrificing computational time. The fundamental difference between EPnP and other pose solvers lies in the idea of writing the coordinates of 3D points as a weighted sum of 4 virtual control points which, then, are expressed in the camera reference frame as linear combinations of a 12x12 matrix eigenvectors, thus reducing the problem to a small constant number of quadratic equations to determine the correct weights. In addition to point-based pose estimation, recent studies have proved the effectiveness of PnL (Perspective-n-Line) algorithms due to the more robust invariancy to lighting conditions of edges and lines. Some example of PnL pose solvers are Absolute Subset-based PnL (ASPnL) which performs better for small lines **[47]**, Linear formulation PnL (LPnL) **[47]**, a linear pose estimation proposed by Ansar and Daniilidis **[48]** and a method proposed by Mirzaei and Roumeliotis **[49]**.

The refinement of the solution is the last step of the pose estimation process and implies local or global optimization. In the case of pose initialization, the result of a first guess calculation can be optimized through the Newton-Raphson Method for iterative minima research. The fit error to minimize is the difference between detected features and projected 3D points. A second local optimization technique that derives from NRM, the Gauss-Newton algorithm, is introduced along with the EPnP solver in order to improve the accuracy of pose estimates by minimizing a sum of squares with non-linear regression: the iterative process is performed only over four

coefficients to minimize the square difference of control points coordinates in the camera reference frame. On a global level, the bundle adjustment (BA) approach refines the results of pose tracking in a series of relative estimates by reducing the reprojection error in each image. BA optimization relies on the Levenberg-Marquardt algorithm to solve linear systems and formulate better values for the camera intrinsic parameters and the scene geometry. Lastly, the implementation of an Extended Kalman Filter (EKF) architecture for relative navigation allows predictive state estimation and noise filtering to contribute to assess the consistency of successive position and attitude measurements.

2.2 Algorithm description

The main goal of this thesis is to provide further validation to a robust monocular vision-based pose initialization algorithm introduced by S. Sharma, J. Ventura, S. D'Amico in "*Robust Model-Based Monocular Pose Initialization for Noncooperative Spacecraft Rendezvous*" **[31]**. Along with the techniques described in the previous paragraph, the so-called SVD architecture suggests interesting innovations to address the problem of pose determination for close-range proximity operations in the scenario of a known target without fiducial markers. The paper describes the algorithm, pointing out the difference with the state-of-the-art techniques and comparing performance results among various configurations of feature extractors, ROI detection and pose solvers to demonstrate the feasibility of the SVD method. The problem of pose initialization consists in computing the rotation matrix $\boldsymbol{R}^{BC}$ and the translational vector $\boldsymbol{t}^{C}$ that describe the transformation between the camera frame $\{\boldsymbol{C}\}$ and the target body frame $\{\boldsymbol{B}\}$**.** Hence, given a point $\boldsymbol{q}^{B}$ of the 3D spacecraft's model whose coordinates $[\,x^{B}, y^{B}, z^{B}]^{T}$ are expressed in the body reference frame, the correspondent point described in the camera frame $\boldsymbol{r}^{C}$, its coordinate and its projection on the image plane $\boldsymbol{p} = [u, v]^{T}$ are obtained by using the following perspective projection equations:

$$\boldsymbol{r}^{C} = [\,x^{C}, y^{C}, z^{C}]^{T} = \boldsymbol{R}^{BC}\boldsymbol{q}^{B} + \boldsymbol{t}^{C} \tag{2.1}$$

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

where f_x and f_y represent the focal lengths of the camera and (C_x, C_y) the principal point of the image. Moreover, the camera frame is oriented as shown in **Fig. 2.1** with the axis C_3 along the boresight,C_2 and C_1 aligned with the image frame $\{P_1, P_2\}$.

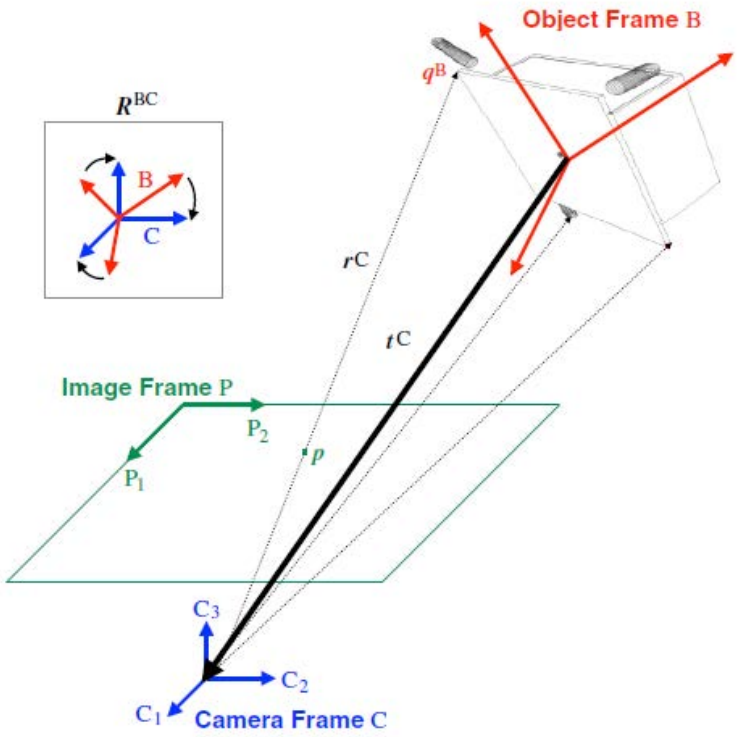


Figure 2.1 Schematic overview of the pose estimation problem with monocular camera **[31]**.

The structure of the proposed architecture is shown in **Fig. 2.2** and an overview of the SVD algorithm and the paper's results is presented in the next paragraphs.

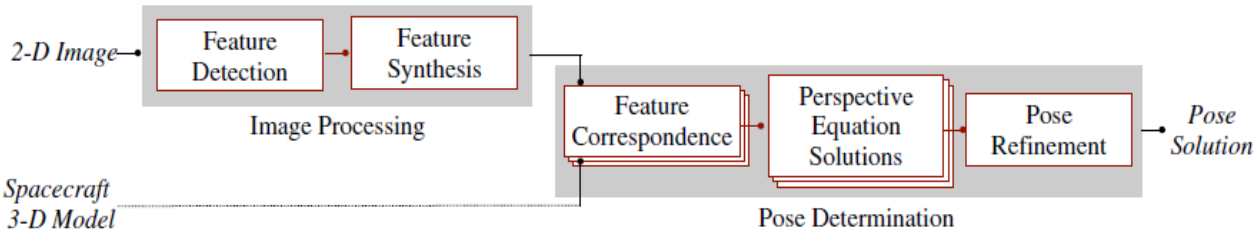


Figure 2.2 Architecture of SVD method for pose estimation **[31]**.

2.2.1 Image Processing

After an initial pre-processing procedure is performed on the raw images to correct them for lens distortion, a Gaussian filter is applied to decrease the magnitude of image noise and the resulting data is fed as input to the Weak Gradient Elimination (WGE) block which is entitled to identify the target's Region of Interest. Based on the hypothesis that, under most illumination conditions, space vehicles present more pronounced gradient variations in correspondence with the structure's edges than the background, whether it's empty or the surface of the planet is visible. The computation of the gradient distribution $\boldsymbol{G}$ is obtained at all pixel locations by using a convolution kernel like the Prewitt operator. As in **Eq. 2.3**, the horizontal and vertical derivative are the product of the convolution operation $*$ between the input image $\boldsymbol{A}(u, v)$ and two 3x3 matrices that approximate the derivative by means of finite differences. The next step consists in thresholding the gradient's magnitude (**Eq. 2.4**) in order to eliminate the weakest elements in the image.

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -1 & 0 & +1 \\ -1 & 0 & +1 \end{bmatrix} * A \qquad G_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ +1 & +1 & +1 \end{bmatrix} * A \tag{2.3}$$

$$G(u, v) = \sqrt{G_x^2(u, v) + G_y^2(u, v)} \tag{2.4}$$

The gradient is normalized, sorted in 100 uniformly distributed bins and fitted by an exponential probability distribution function (PDF), then, all the pixel which correspond to a cumulative distribution inferior to 0.99 are classified as "weak". Once the gradient of weak pixels is set to zero, only the most prominent features in the image are left to estimate a ROI. By setting limits to the cumulative distribution function (CDF) of the gradient along the vertical and horizontal axes it's possible to identify a rectangular region of interest where the features that yield the strongest intensity variations are located. For instance, the ROI limits can be set between 0.025 and 0.975 to enclose the central 95% of the gradient if a normal distribution is assumed.

The determination of a target's ROI combined with the knowledge of the spacecraft 3D model allows a coarse estimation of relative range and LOS. Given the ROI's diagonal length l_{ROI} and the characteristic length L_C of the spacecraft and the focal lengths of the camera f_x and f_y, the range between camera frame and target is defined by **Eq. 2.5 .** Moreover, assuming (C_x, C_y) as the principal point of the image and

calculating the ROI's center (B_x, B_y) the azimuth and elevation angles (α, β) are defined by **Eq. 2.6**.

$$\|t^C\|_2 = \frac{((f_x+f_y)/2)\cdot L_C}{l_{ROI}} \tag{2.5}$$

$$\alpha = \tan^{-1}\left(\frac{B_x - C_x}{f_x}\right) \qquad \beta = \tan^{-1}\left(\frac{B_y - C_y}{f_y}\right) \tag{2.6}$$

Lastly, the relative position vector is calculated as follows:

$$t^C = \begin{bmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & \sin\beta \\ 0 & -\sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \|t^C\|_2 \end{bmatrix} \tag{2.7}$$

Thereafter, the SVD image processing algorithm employs a feature extraction architecture based on merging two separate streams of data in order to detect different elements of the image and provide more robustness to the results. The authors opt for edge detection techniques since, as previously explained, linear features show a less variable response to illumination. Besides, by pre-emptively determining a ROI, edge detectors' hyperparameters can be automatically tuned to operate on a narrower range of possibilities and produce more accurate data to feed to the feature matching algorithm. The first stream of features is extrapolated from the filtered gradient image by means of a Hough transform. The Hough transform performs a shape identification through a voting procedure in a parameter space. In particular, lines are parametrized in polar coordinates $[\rho, \theta]$ and the voting procedure selects a series of candidate linear objects on the basis of two fundamental parameters: the expected minimum length of the segments l_{min}^H and the maximum gap between points to be considered points of the same line λ^H (**Eq. 2.8**). The choice of correct parameters greatly affects the effectiveness of the feature extraction phase both in terms of precision and predictivity. To improve the flexibility of the edge detection without manually tuning the Hough transform for every image the following relations are imposed:

$$l_{min}^H = k_1 \cdot l_{ROI} \qquad \lambda^H = k_2 \cdot l_{ROI} \tag{2.8}$$

where k_1 and k_2 are numerical coefficients that can be empirically estimated through iterative offline testing with simulated datasets and with real images from previous close-range approaches.

A second stream of features is obtained with the application of the Sobel operator to the undistorted unfiltered image. The Sobel operator, much like the Prewitt operator, employs matrix convolution (**Eq. 2.9**) to filter and differentiate the intensity distribution of the image along the horizontal and vertical axes.

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * A \qquad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * A \tag{2.9}$$

Once the edges in the image are identified the Hough transform is implemented again to identify the lines that belong to the target vehicle and the hyperparameters l_{min}^{H} and λ^{H} are adaptively computed as scalar multiples of l_{ROI} by the coefficients k_3 and k_4. Since the Hough transform is applied to the entirety of the image, only the segments whose midpoint lie inside the ROI's limits are considered valid.

Moreover, further feature simplification is suggested by Sharma et al. **[31]** to reduce the fragmentation of segments that result from Hough line detection algorithm. As shown in **Fig. 2.3**, the presence of similar adjacent lines indicates the possibility of a truncated edge and merging the segments can effectively reduce the uncertainty and computational cost of feature matching and pose solving. Hence, certain geometrical conditions must be defined to check if two segments can be considered similar and merged in a single edge. The paper proposes as conditions of similarity three thresholds θ_{th}, ρ_{th}, d_{th} respectively for orientation, radius and distance between the farthest endpoint with whom the two groups of detected edges must be tested.

Finally, the feature detection phase is concluded by merging the two streams of features and combining the results. The aim of combining two different edge detection procedures is to identify different linear elements of the spacecraft's structure: the first architecture (WGE + Hough) is dedicated to detecting the smaller features in the image such as antennas, the second one (Sobel + Hough) is dedicated to recognize larger edges instead. The process of merging the streams is carried on by testing the similarity of the lines in terms of orientation, radius and distance between midpoints. Also, the thresholds that define the condition of similarity can be expressed as functions of l_{ROI} to make them independent from the relative range. If two lines are classified as similar, only the longer one is retained and in the case of

intersections, assuming the shorter line is divided in two portions (L1 and L2), only the longer section is preserved if $L_1/L_2 > 0.25$.

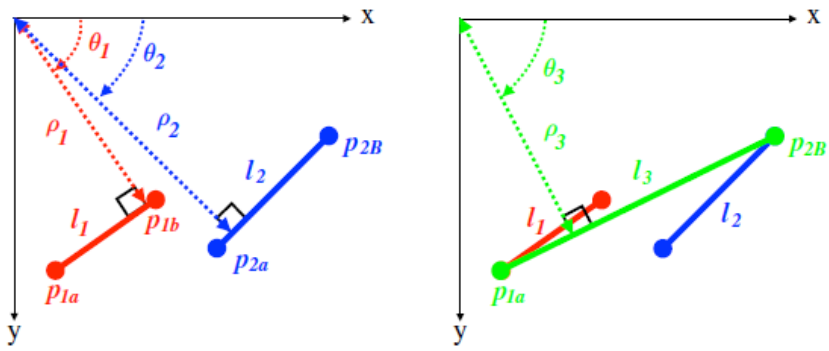


Figure 2.3 Merging of two truncated edges **[31]**.

2.2.2 Feature Synthesis and Matching

The main innovation introduced in the SVD architecture is the perceptual grouping method to reduce the number of associable features. Perceptual grouping is the process of synthetize all the edges detected in the image into higher-level features by checking the validity of geometric relations among the segments. To resolve uniquely the pose problem (PnP) a minimum number of six corresponding 2D and 3D points are required, thus, the total amount of combinations that compose the search space for feature matching is equivalent to:

$$\binom{m}{6}\binom{n}{6}6! \tag{2.10}$$

where *n* and *m* represent respectively the number of image points and world points. Nonetheless, the choice of adopting perceptual grouping is based on the assumption that the target's structure model is sufficiently simple to be decomposed into a subset of more or less complex geometric elements.

The edges are divided into six high-level groups which are, in order of decreasing complexity:

1. Polygonal tetrads
2. Polygonal triads
3. Parallel triads

4. Parallel pairs
5. Proximity pairs
6. Antennas

The process of feature selection and synthetization is carried on by examining a few geometric constraints regarding orientation, distance and length. In fact, two lines are identified as a parallel pair if the relative difference in orientation lies within a tolerance angle θ_{max} or as a proximity pair if the distance between the closest endpoints is shorter than the threshold d_{max}. More complex elements are built by association of simpler ones: two parallel pairs with a shared segment compose a parallel triad as well as two proximity pairs make up a polygonal triad and, lastly, two polygonal triads with two edges in common are classified as a closed polygonal tetrad. Separately, edges with a length shorter than l_{max} are categorized as antennas. A schematic representation of the classes of high-level features are displayed in **Fig. 2.4**. Every condition that is established to determine the geometric relations among the detected lines can be adaptively tuned as function of l_{ROI} and in relation to the actual proportions of the spacecraft's model to improve the algorithm's robustness to range variations. The importance of defining variably complex elements is related to the possibility of characterizing accurately both easily identifiable simpler features such as antennas and more complex prominent features like solar panel surfaces at

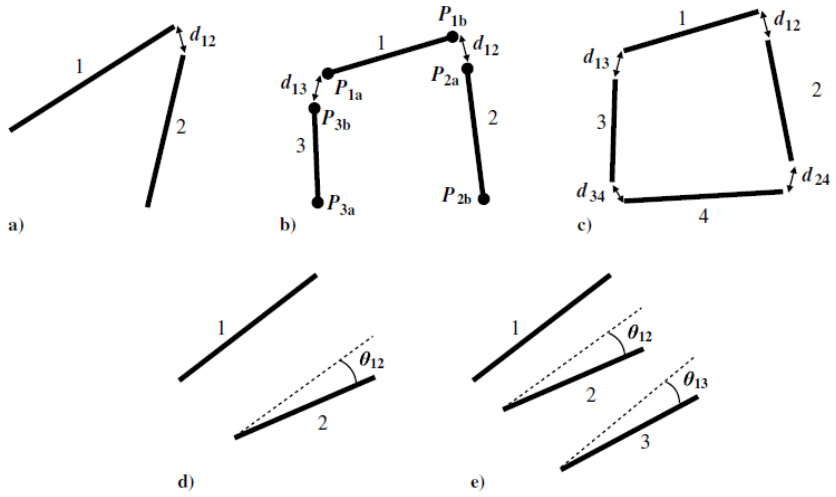


Figure 2.4 High-level feature groups: proximity pair (a), polygonal triad (b), polygonal tetrad (c), parallel pair (d), parallel triad (e) **[31]**.

the same time. Similarly to the image lines, a perceptual grouping technique is applied to a wireframe model of the target vehicle in order to obtain the same 3D high-level features.

The feature correspondence algorithm proposed by Sharma et al. **[31]**, unlike previous view-based pose estimation architecture suggest, doesn't take into consideration all the possible combinations of 3D and 2D elements in the creation of a match matrix. Since most of the high-level features don't provide enough points to obtain a unique solution to the PnP problem, only a combination of two of them is required to feed as input to the pose solver. The authors propose to achieve an effective solution by combining the most complex and the simplest feature groups detected, in particular antennas are usually easily identifiable, if present, while closed polygonal tetrad like solar panels or large flat surfaces are less prone to the probability of accidental detection **[50]**. Moreover, the number of features is inversely proportional to the geometric complexity and this solution drastically reduces the search space for pose solvers. All the possible matching combinations are stored in a so-called match matrix.

2.2.3 Pose solution and refinement

The EPnP pose solver is the technique chosen to determine the rotation matrix $\boldsymbol{R}^{BC}$ and the translational vector $\boldsymbol{t}^{C}$. As described previously, EPnP can provide efficient, robust and fast pose estimation for a larger number *n* of features by expressing the points as a linear combination of four noncoplanar unknown control points $\boldsymbol{c}_1^C, \boldsymbol{c}_2^C, \boldsymbol{c}_3^C, \boldsymbol{c}_4^C$ as in **Eq. 2.11**.

$$\boldsymbol{r}_i^C = \sum_{j=1}^{4} \gamma_{ij} \boldsymbol{c}_j^C \qquad i = 1,2, \ldots, \tag{2.11}$$

where $\boldsymbol{r}_i^C = [\, x^C, y^C, z^C]^T$ are the 3D points expressed in the camera frame.

Assuming the validity of the perspective projection equation (Eq. 2.2) for the pinhole camera model, the PnP problem is solved by determining the 12 unknown coefficients for the control points in a linear system of $2 \times n$ equations. For each pair of features the equations are the following:

$$\sum_{j=1}^{4} \gamma_{ij} f_x \boldsymbol{c}_{x_j}^C + \sum_{j=1}^{4} (C_x - u_i) \gamma_{ij} \boldsymbol{c}_{z_j}^C = 0 \qquad i = 1,2, \ldots, n$$

$$\sum_{j=1}^{4} \gamma_{ij} f_y \boldsymbol{c}_{y_j}^C + \sum_{j=1}^{4} (C_y - v_i) \gamma_{ij} \boldsymbol{c}_{z_j}^C = 0 \qquad i = 1,2,\dots,n$$

(2.12)

Finally, a cumulative reprojection error is defined in **Eq. 2.13** to operate as a confidence metric and evaluate the accuracy of the pose solver algorithm.

$$E_{2D} = \frac{1}{n} \sum_{i=1}^{n} \sqrt{\left[u_i - \left(\frac{x_i^C}{z_i^C} f_x + C_x\right)\right]^2 + \left[v_i - \left(\frac{y_i^C}{z_i^C} f_y + C_y\right)\right]^2}$$

(2.13)

In regard to the optimization method the authors suggest a Newton-Raphson for the five combinations of features with the lowest reprojection error to further refine the pose estimates and select the optimal one. A fit error vector $\boldsymbol{E}_S = [E_1 \dots E_n]^T$ and the Jacobian matrix [$\boldsymbol{J}$] are calculated at each step of the iterative process and are used to update the state variable $\boldsymbol{x}$ which contains six unknown parameters, e.g., three Euler angles for the rotation matrix θ_{BC} and three coordinates for the translation vector t_C .

$$E_i = \left[u_i - \left(\frac{x_i^C}{z_i^C} f_x + C_x\right), v_i - \left(\frac{y_i^C}{z_i^C} f_y + C_y\right)\right] \quad i = 1,2,\dots,n$$

(2.14)

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - (\boldsymbol{J}^T \boldsymbol{J})^{-1} \boldsymbol{J}^T \boldsymbol{E}_S(\boldsymbol{x}_k)$$

(2.15)

2.2.4 Preliminary validation

The proposed SVD architecture is validated on PRISMA mission's data collected by the onboard optical sensor (VBS). As discussed in the previous chapter, the PRISMA mission consisted in two separate spacecrafts, MAIN (Mango) and TARGET(Tango), with different attitude control ability, that emulate typical scenarios of partially or totally noncooperative proximity operations. The validation is conducted on a set of real space imagery that portray the Tango vehicle at slightly different ranges and orientations with various background. A wireframe model of the target provides a description of the basic geometric relations among the edges and their coordinates in order to extract the 3D features and organize them in the aforementioned feature

groups. A body-fixed reference frame is defined at the center of the bottom surface of the vehicle's body as it's displayed in Fig. 2.5. The authors define 16 proximal pairs, 18 parallel pairs, 12 parallel triads, 6 closed polygonal tetrads, 6 open polygonal tetrads, and 5 antennas as 3D features to combine with the elements detected in the image. True reference for the relative range estimation is obtained by the Phoenix GPS receiver, whereas sun-sensors and magnetometers produce coarse attitude measurements, thus reaching an accuracy of 2 cm and 3° (3D rms) [14] for the "true" relative pose ($\boldsymbol{R}_{true}^{BC}$, $\boldsymbol{t}_{true}^{C}$).The translation error is evaluated as absolute difference between the translation vectors and the rotation error is described by three Euler angles which describe the rotations around the camera frame axes. Other parameters to measure the performances of the SVD architecture employed in the paper are the 2D reprojection error, the true positive rate TPR and the positive predictive value PPV which evaluate respectively the pose solver and the feature detection effectiveness.

$$TPR = \frac{True\ positives}{True\ positives + False\ positives} \cdot 100 \tag{2.16}$$

$$PPV = \frac{True\ positives}{True\ positives + True\ negative} \cdot 100 \tag{2.17}$$

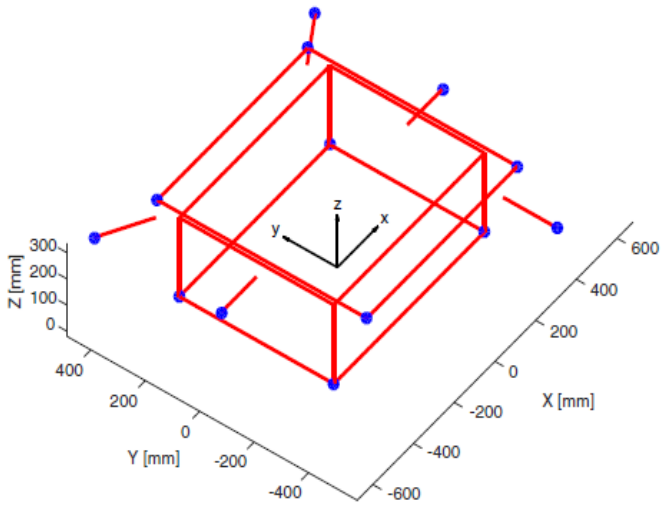


Figure 2.5 Tango's 3D wireframe model. **[31]**

The first test revolves around the choice of the optimal feature extractors. The authors compare on a set of 142 images of Tango, the PPV and TPR of state-of-the-art

techniques with the WGE algorithm proposed in the paper. The feature detectors tested are a combination of Sobel, Canny and Prewitt with the Hough transform, the features from accelerated segment test (FAST) technique and a key-point detector like BRISK. If the detected edge's endpoint is located within an Euclidean distance of 5 pixel from a true edge endpoint it's classified as true positive.

As shown in **Fig. 2.6**, the WGE architecture has a higher TPR (32.8%) than other feature detectors, even though it lacks location precision in respect to Canny and Sobel. The performance of WGE are comparable to those of FAST and BRISK in terms of positive predictive value, yet the computational time required by WGE is an order of magnitude greater. Besides, the WGE algorithm is ten times faster than the other edge detectors (Sobel, Canny, …) since it is applied to a more limited image area and the search space is reduced, therefore it has better performances on average.

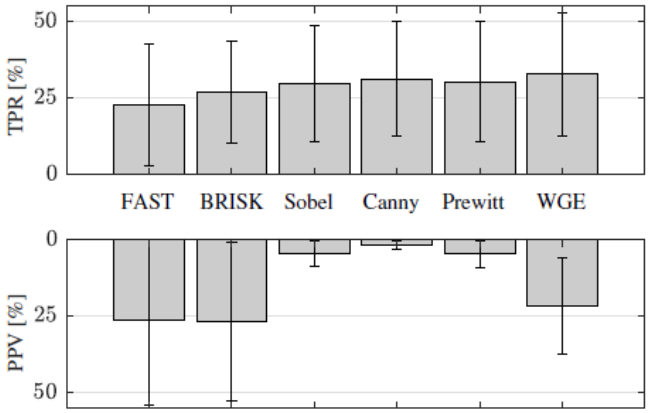


Figure 2.6 Mean and standard deviation of PPV and TPR for different feature extractors. **[31]**

Secondly, the WGE algorithm has been tested for ROI detection in comparison to the MSER blob detector. MSER allows to define a list of bounding boxes which, subsequently, are fed to a non-maximum suppression (NMS) algorithm and confined in a single region of interest. The parameters PPV and TPR are defined in relation to the true rectangular box that encloses the projected points of the reference true pose. The results on the PRISMA dataset demonstrate that the two method are comparable in terms of effectiveness, i.e., mean PPV = 85.98% and mean TPR=90.59% for WGE and, respectively, 81,77% and 89.45% for MSER+NMS. On the other hand, the weak gradient elimination technique is performed in a mean computational time of 0.0878 s, five time faster than MSER, and can't identify multiple ROIs.

A third test is performed to evaluate the accuracy of different pose solvers. In **Fig. 2.7** the success rate of multiple PnL algorithm and EPnP is plotted in relation to the number of line correspondences, showing clearly how EPnP can provide a sufficiently high percentage of accurate estimates even with the lowest number of features. Instead, by using a dataset of five synthetic images and simulating variations in image noise with a Gaussian filter, the paper proves that the accuracy of EPnP as pose solver remains superior to the one of PnL algorithms independently from the measurements' noise. It should be noted that a pose estimate is considered successful if $\|E_T\|_2 < 30\ cm$ and $\|E_R\|_2 < 10°$.

The EPnP architecture proves to be the best solution to the pose estimation problem in terms of robustness and success rate, yet the pose outputs can be greatly affected by the results of the feature synthesis segment of the SVD architecture. A final test, based on a dataset of 25 images, is performed to compare the EPnP algorithm with a RANSAC approach. The process of applying random matching hypotheses to the detected points and selecting the larger number of inliers that produce the lowest reprojection error (RANSAC) provides a pose solution for all the images, yet it doesn't reach the same level of accuracy that the SVD algorithm has in the case of high confidence estimates. As reported in **Tab. 2.1**, the SVD method determines successfully the relative pose only in a small percentage of the images with a sufficient rotational and translational precision, whereas the RANSAC approach gives generally lower accuracy in every pose result. The main limitation of both methods is the wrong edge detection and grouping which occurs with partial edges, duplicated lines and elements that are not present in the 3D model. Nonetheless, RANSAC requires on average a longer computational time than SVD.

SVD solution type	Number of images	Method	E_R, deg	E_T, m
High-confidence pose	5	SVD	(−0.57, 0.59, −1.37)	(0.14, 0.06, 0.51)
		RANSAC	(29.84, 7.52, −17.87)	(0.46, 0.52, 1.94)
Low-confidence pose	7	SVD	(−23.56, −0.67, 16.78)	(0.18, 0.005, 0.75)
		RANSAC	(75.66, 2.18, −22.16)	(0.28, 0.35, 1.13)
Relative position only	13	SVD	— —	(0.07, 0.03, 0.51)
		RANSAC	(−10.04, −2.11, 23.41)	(0.24, 0.48, 1.38)

Table 2.1 Average errors of pose estimates with SVD and RANSAC **[31]**.

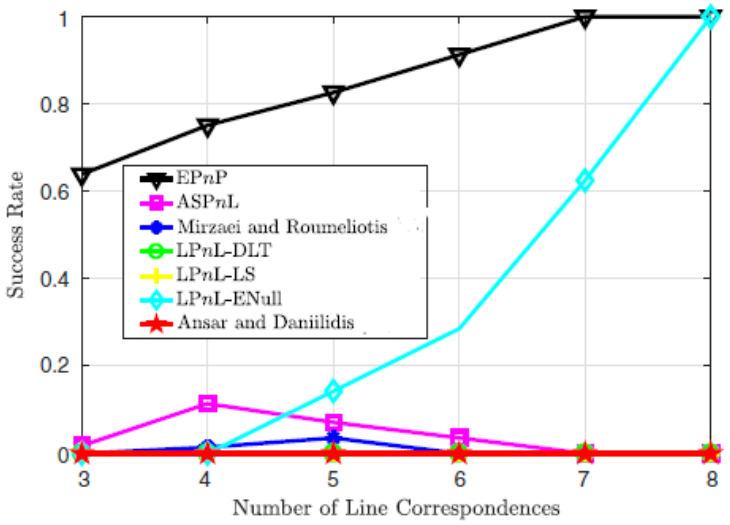


Figure 2.7 Average success rate as a function of line correspondences for different pose solvers **[31]**.

Chapter 3

Algorithm validation on PRISMA synthetic imagery

The chapter contains the results of a validation test of the SVD pose initialization technique based on a set of synthetic imagery of the Tango spacecraft. A report of the specificities of the algorithm implemented in the thesis work is presented. Moreover, the second part of the chapter is focused on the analysis of the results of the validation in terms of accuracy, effectiveness and success rate of the pose estimator.

3.1 Algorithm implementation

The thesis revolves around the implementation and further validation of the SVD architecture for pose initialization described in the previous chapter. By following the indications contained in **[31]**, a similar version of the algorithm is implemented within the MATLAB programming environment and tested on a larger database of synthetic images of the PRISMA target vehicle, Tango. The purpose of providing further validation consists in determining the feasibility and the range of applicability of this pose estimation technique and highlighting the main disadvantages. Moreover, the thesis focuses on proposing some possible modifications to the algorithm that would improve the perceptual grouping process. Since the aforementioned paper only describes some preliminary results for the SVD architecture, providing experimental proof for the choice of the ROI detection subsystem, the feature extractor and the pose solver, the validation test described in the thesis hinges on the evaluation of the method's performance in a wider range of conditions. The testing is performed to assess the effectiveness of pose estimation in relation to various factors: presence of composite background, variation in relative distance, number of feature correspondences.

3.1.1 The SPEED dataset

The validation test is executed on the Spacecraft Pose Estimation Dataset (SPEED), a set of 12000 computer-generated images of the Tango vehicle. As described in **[51]**, SPEED is available to the scientific community to perform comparisons and testing among state-of-the-art pose estimation methods, the dataset has been released publicly in collaboration with the European Space Agency within the KELVINS competition **[52]**. At the Space Rendezvous Laboratory (SLAB) at Stanford University the images have been generated by using two integrated sources, the Optical Stimulator software and real images of the Earth inherited by the Himawari-8 geostationary mission observation data. Optical Simulator emulates camera properties and renders synthetic images using MATLAB and C++ bindings of OpenGL. A randomly uniform distribution of relative positions and attitudes are employed to produce diverse views of the target vehicle and improve the quality of the dataset. In fact, a random set of rotations, parametrized as unit quaternions, translations and bearing angles is calculated by means of a standard normal distribution. Relative distances outside the range between 3 m and 50 m are rejected.

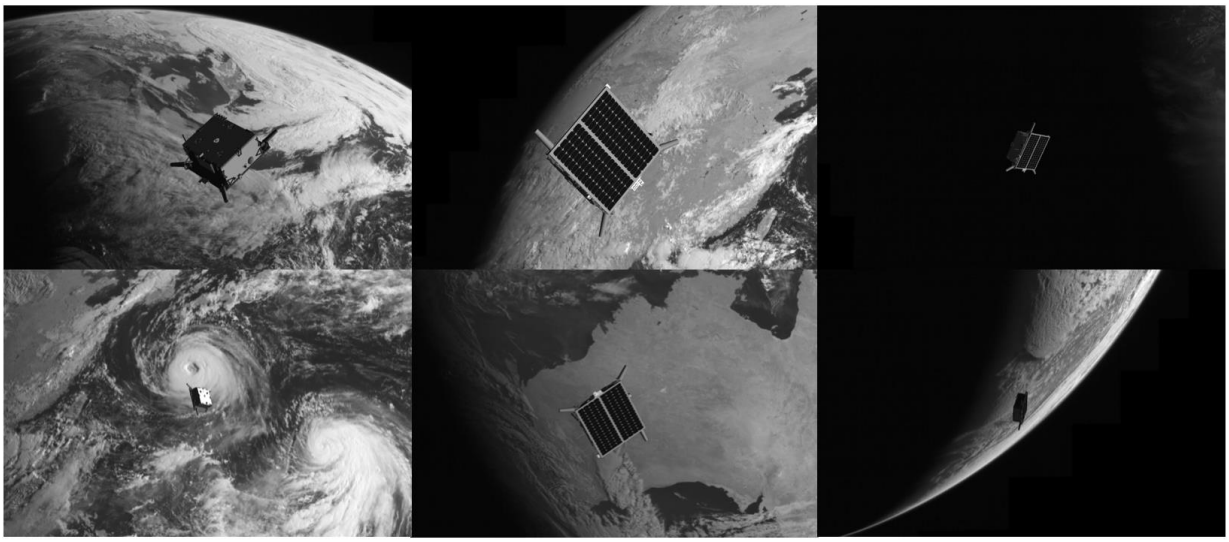


Figure 3.1 A montage of some synthetic images with computer-generated background from the SPEED dataset. **[51]**

The background is generated from a series of 72 full-disk Earth actual images, the data has been captured by a geostationary meteorological system. The images have a $100 \cdot 10^6$ pixel resolution, taken over a period of 12 hours and, in order to match the scale of the Earth when viewed through the camera, they are cropped and randomly distributed among half of the synthetic images. Azimuth and elevation angles for the solar illumination are chosen to be combined with illumination of the 72 Earth

pictures. The scale of the cropped sections is adjusted to meet the hypothesis of a chaser orbiting at an altitude of 700 km and in a nadir-pointing attitude. Lastly, the background is blurred by means of Gaussian smoothing function and white noise is added to simulate, respectively, the depth of field an shot noise. Some examples of the images from the SPEED database are presented in **Fig. 3.1**.

The Optical Simulator consists in a virtual reality testbed with a pair of lenses that magnify a monitor. High-fidelity, synthetic undistorted scenes of the space environment are rendered to the monitor in real-time and closed-loop to stimulate a vision-based sensor test article **[52]**. The camera employed is a Point Grey Grasshopper 3, equipped with a Xenoplan 1.9/17 mm lens. The intrinsic properties of the camera are described in **Tab. 3.1**.

Parameter	Description	Value
N_u	Number of horizontal pixels	1920
N_v	Number of vertical pixels	1200
f_x	Horizontal focal length	17.6 mm
f_y	Vertical focal length	17.6 mm
d_u	Horizontal pixel length	$5.86 \cdot 10^{-3}$ mm
d_v	Vertical pixel length	$5.86 \cdot 10^{-3}$ mm

Table 3.1 Parameters of the camera used to capture the SPEED images.

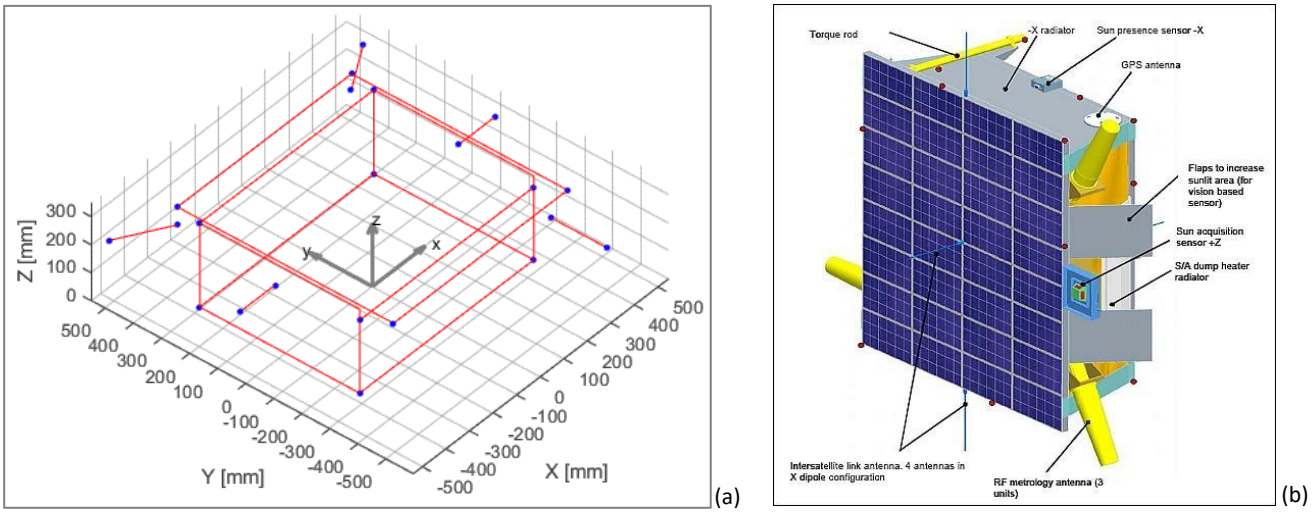
3.1.2 Tango's spacecraft model

A digital structure of Tango spacecraft is reduced to a wireframe model to highlight the linear elements of the vehicle that can be easily identified. In **Fig. 3.2** the computer-generated model and its wireframe equivalent are displayed. The spacecraft fundamentally consists in a solar panel (560×750 mm), a convex polyhedron representing the vehicle body ($560 \times 550 \times 300$ mm) and five additional segments, representing the RF metrology antennas and intersatellite communication antennas, with a length of 204 mm. The body-fixed reference frame for the target is positioned at the center of the body bottom surface. Within the wireframe Tango model, the following number of high-level features can be detected:

- 6 polygonal tetrads

- 6 polygonal triads
- 12 parallel triads
- 18 parallel pairs
- 16 proximal pairs
- 5 antennas

The 3D geometric features and the 3D point coordinates that compose the target model are, respectively, defined and computed by applying the same geometric conditions employed for the 2D features.



(a) (b)

Figure 3.2 A wireframe model (a) and a computer-generated model of PRISMA's TARGET satellite, Tango.

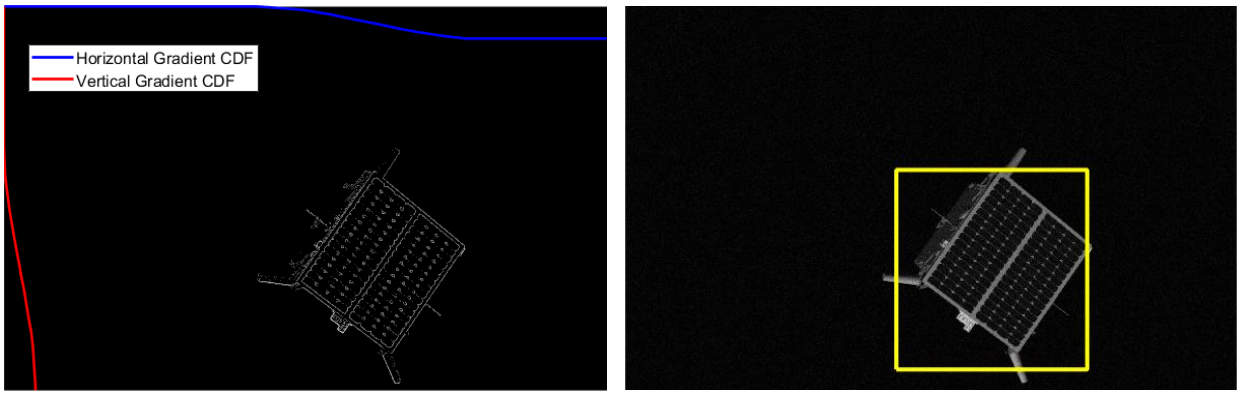
3.1.3 The image processing subsystem

The first block of the pose estimation technique is the image processing subsystem. Each image is subjected to a preventive gaussian filtering with the built-in function *imgaussfilt*, then it's fed as input to the WGE block. Firstly, the image gradient is calculated by means of Prewitt operator, secondly the gradient distribution is normalized and fitted with an exponential distribution (**Eq. 3.1**). The distribution is limited to the the 1% strongest gradient pixels with a threshold and the remaining points are set to zero. Thereafter, vertical and horizontal gradient are determined and the respective cumulative distribution functions (CDFs) are computed. After selecting a limit percentage for the procedure of ROI detection, such as the central 95 % of the distribution, the coordinates of a bounding box are selected.

$$y = \frac{1}{\mu} e^{-\frac{x}{\mu}} \tag{3.1}$$

In the case of the SPEED data, since half of the images present composite background and the detection of a correct ROI is generally limited by the presence of interference, the best choice for a gradient range threshold is conservative, in fact the bounding box limits are set to 0.05 and 0.95 and, therefore, only the central 90% of the CDF is selected. However, restricting the range of detection affects negatively the images where the background is empty and for this reason an additional constant, equal to 5% of the mean edge length of the ROI, is employed to automatically extend the region's boundaries. In this way the stricter threshold for gradient limits is balanced by the constant offset in the case of correct ROI identification.

An example of the effectiveness of the algorithm is shown in **Fig. 3.3**.



(a) (b)

Figure 3.3 The cumulative distribution function of the gradient along the image axes (a) and the results of the ROI detection process (b).

The consecutive section of the algorithm offers a first coarse evaluation of the relative distance and line of sight angles assuming that the ROI correctly circumscribes the target's figure in the image. The results about the accuracy of the estimations are discussed in the next paragraphs. Furthermore, the parameter l_{ROI}, length of the ROI's diagonal is calculated.

A first edge detection process is applied to the gradient image by means of the Hough transform. Before the application of the transform the gradient outside the region of interest is set to zero to make sure that it is only implemented within those limits. In MATLAB the Hough transform requires the definition of some hyperparameters such

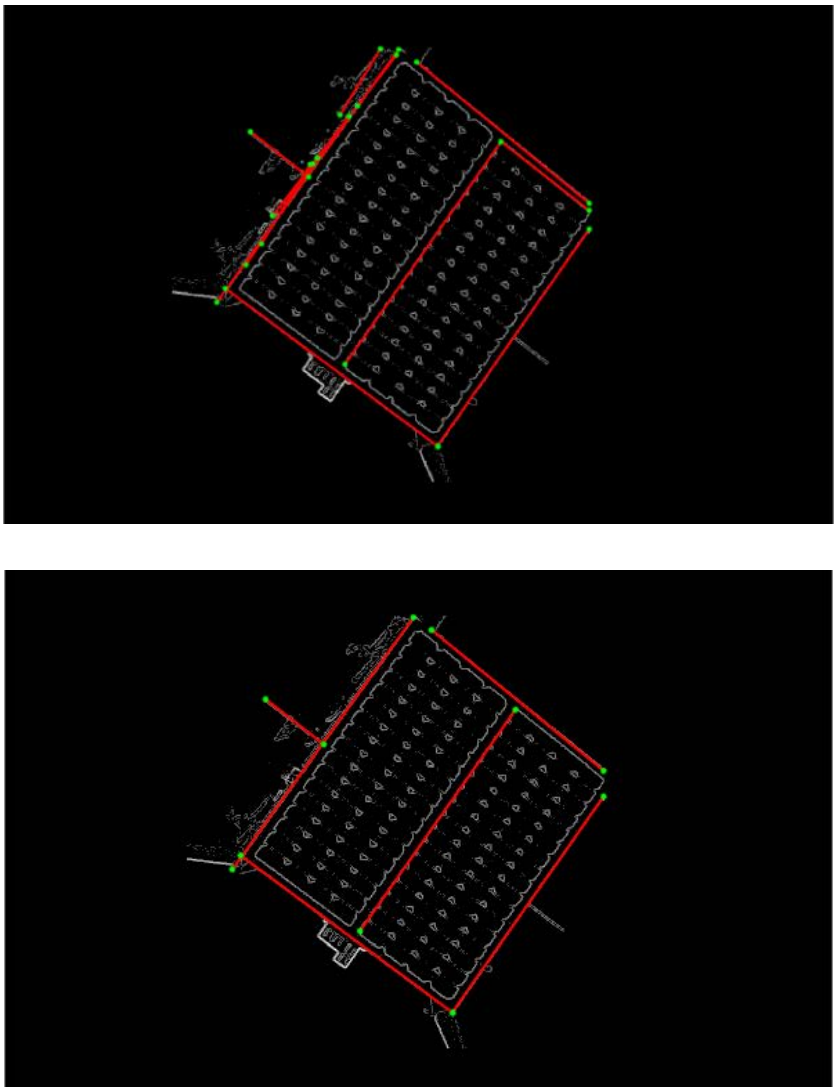
as the resolution of ρ and θ, the maximum number of peaks and a minimum threshold for their value, while the main geometric constraints that can be defined are the minimum length of the line segments l_{min}^{H} and the maximum gap between points λ^{H}. As set in the proposed algorithm, the lines are identified with a radial resolution of 0.5 pixel and an angular accuracy of 0.1° over the range [-90 89.9], once the data retrieved from the image are represented in polar coordinates a maximum number of 15 peaks over a threshold of 0.1. The geometric conditions, as suggested in the paper **[31]**, are defined as scalar multiples of l_{ROI}, which acts as a fiducial metrics for the scalability of those parameters. The parameters are the following:

$$l_{min}^{H} = 0.075 \cdot l_{ROI} \qquad \lambda^{H} = 0.01 \cdot l_{ROI} \tag{3.2}$$

Consecutively, a section of the algorithm is dedicated to filtering fragmented lines and duplicated edges. The authors suggest imposing a similarity condition by means of an angular and radial threshold equal to the resolution of the Hough transform, instead the necessity for larger limits emerges. In **Fig. 3.4** a comparison between the result of a selection process with accuracy of 0.5 px and 0.1° and the results when an angular limit of 10° and a radial limit equal to $0.1 \cdot l_{ROI}$ is imposed. Clearly larger geometric conditions can improve effectively the identification of the spacecraft's true edges. Moreover, a distance threshold equal to half the length of the longer line is set to be checked between the midpoints of every pair of segments that respect the similarity conditions, given the fact that some antennas can be regarded as duplicated edges and even though they aren't.

As explained in the previous chapter, the stream of data processed by the Hough transform is coupled with a second stream to improve robustness. Hence, the image is analysed with the MATLAB function *edge* which applies the Sobel operator to the undistorted image. The convolution of the image with the operator produces a binary image with the location of the detected edges, assuming an intensity threshold equal to 0.08. Further filtering is required to eliminate the smaller reflective elements and it is provided by the function *bwareaopen* which removes all the connected components (objects) with less than 10 pixels in the binary image. The Hough transform is subsequently applied to the totality of the image with the same quality parameters of resolution and intensity as the first stream of data, but slightly different geometric constraints, i.e.:

$$l_{min}^{H} = 0.08 \cdot l_{ROI} \qquad \lambda^{H} = 0.02 \cdot l_{ROI} \tag{3.3}$$



(a)

(b)

Figure 3.4 The results of the merging process with different geometric thresholds for orientation and radial distance (a,b).

By setting similar hyperparameters, the two alternative processes for edge detection are expected to identify similar edges, both the longer and contiguous segments of the main structure and the shorter singular antennas, contrary to what the authors propose in **[31]**, namely, the ROI + Hough procedure would be focused on detecting antennas and the Sobel + Hough procedure on detecting longer segments. In this way it's possible to ensure replicability and robustness to all the linear features.

Firstly, the lines which midpoints are located outside the ROI are rejected, secondly, the remaining ones are subjected to a preliminary merging process that unites pairs

of lines that fall within the resolution of the Hough transform (0.1° and 0.5 px). Finally, an additional filtering eliminates overlapped shorter lines with the following similarity conditions:

$$|\theta_1 - \theta_2| \leq 10°$$

$$|\rho_1 - \rho_2| \leq 0.1 \cdot l_{ROI}$$

$$d_{mid} \leq 0.5 \cdot \max(L_1, L_2) \tag{3.4}$$

The results of the Sobel + Hough edge detection are shown in **Fig. 3.5**.

The operation of merging the two streams of data is performed with a dedicated script by defining similar geometric conditions to the ones mentioned before. Initially, among the edges that intersect and overlap only the longer segments are retained, thereafter the lines with an absolute difference of 5° in orientation and $0.05 \cdot l_{ROI}$ in radial distance are merged together. A supplementary threshold for the distance between the farthest endpoints is set in relation to the sum of the lines' dimensions. The results of the combination of streams of features are shown in **Fig. 3.6**.

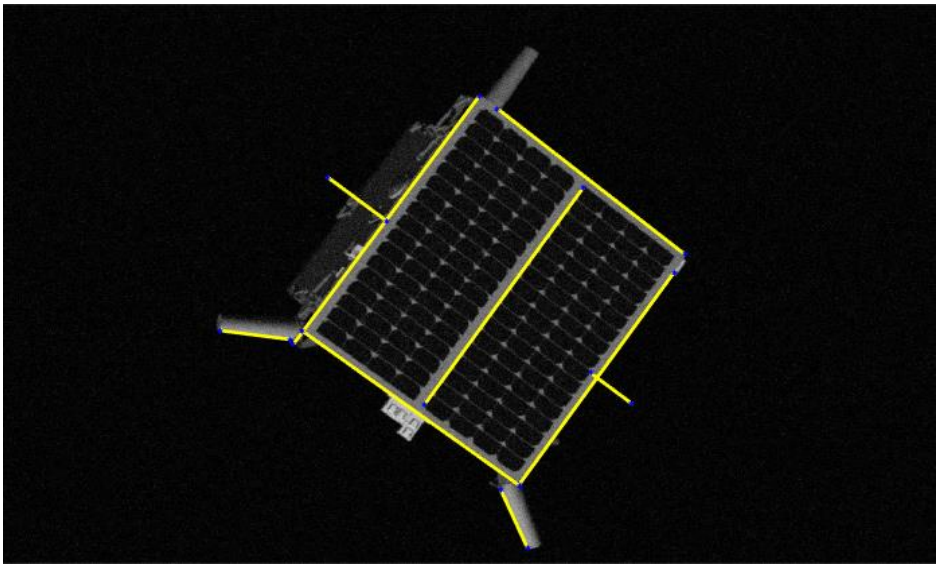
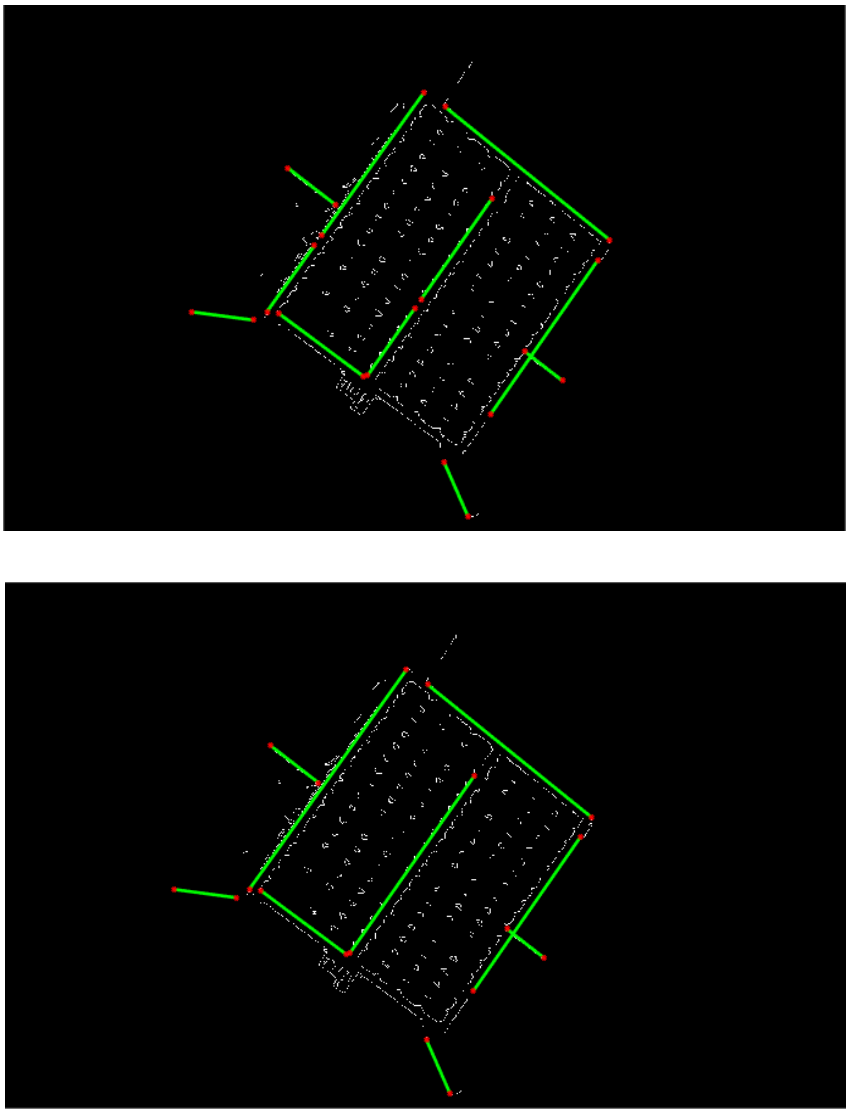


Figure 3.6 The edges resulting from the image processing and feature extraction subsystem.



(a)

(b)

Figure 3.5 The edges detected by Sobel+Hough before the merging process (a) and after (b).

3.1.4 Feature synthesis and matching

The feature synthesis subsystem executes the perceptual grouping task within the SVD algorithm implementation. The function follows the indications of the paper by Sharma et al., yet it introduces some dissimilarities. The proximal pairs are identified by thresholding the distance between the closest endpoints with a superior limit d_{max} and, in addition, a minimum angular difference θ_{min} between the lines is introduced to prevent duplicated edges to be grouped as proximal pair. The following parameters are set:

$$d_{max} = 0.05 \cdot l_{ROI} \qquad \theta_{min} = 10° \tag{3.5}$$

Parallel pairs are recognized by checking the absolute difference of orientation angles among the detected edges, therefore, a maximum angular gap θ_{max} is required. To avoid spurious parallel pairs, it's helpful to assign other two geometric conditions to the radial distance (ρ_{min}) and the minimum ratio between the lines' lengths ($\varepsilon_{min} = L_{min}/L_{max}$). In fact, given the knowledge of the target's structure, it's clearly notable that all the 3D parallel segments have equal length and aren't adjacent, i.e., they don't belong to a common line. The parameters are set as follows:

$$\theta_{max} = 10° \qquad \rho_{min} = 0.05 \cdot l_{ROI} \qquad \varepsilon_{min} = 0.7 \tag{3.6}$$

In regard to antennas, they are selected among the edges that haven't been associated with the group of parallel pairs. A simple criterion is employed to distinguish the antennas from other feature: a maximum length equal to $0.15 \cdot l_{ROI}$ is imposed as threshold.

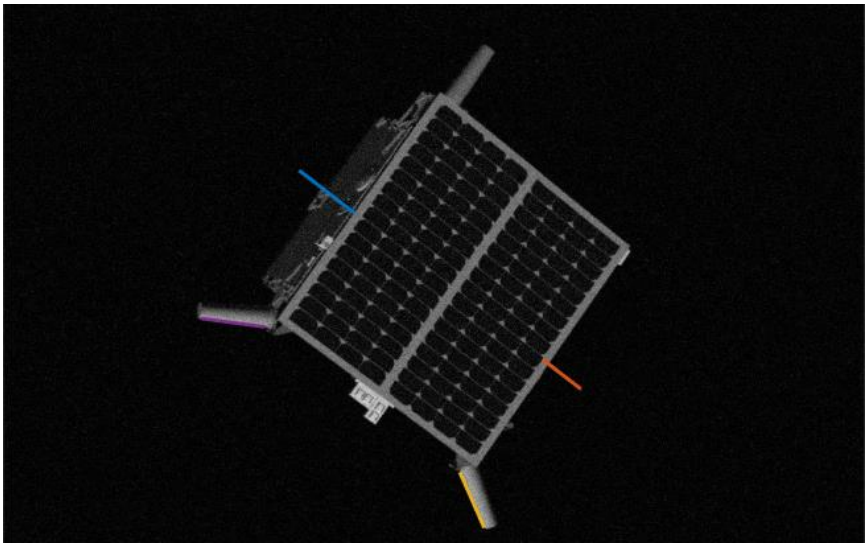


Figure 3.7 The result of antenna's detection.

Parallel triads are individuated by combining parallel pairs with as shared segment, as well as polygonal triads are obtained by coupling consecutive proximal pairs. In particular, open polygonal triads are tested to determine that the endpoints of the non-common lines lie on the same side in respect to the central shared segment. This is ensured by verifying that the line that passes through the two outer endpoints of the triad doesn't intersect the segment that unites their projections on the shared line. As in **Fig. 3.8**, assuming the segments $\{1, 2, 3, 4\}$ compose the proximity pairs

$\{(1,2),(1,3),(1,4)\}$, they can be grouped in 3 polygonal triads $\{(2,1,3),(3,1,4),(2,1,4)\}$, yet by checking the intersection between the lines $(P_{4A} - P_{3B})$, $(P_{4A} - P_{2B})$, $(P_{2B} - P_{3B})$ and the respective segments between their endpoints' projections $(P'_{4A} - P'_{3B})$, $(P'_{4A} - P'_{2B})$, $(P'_{2B} - P'_{3B})$, only one combination is correctly identified as polygonal triad, i.e., the group $(3,1,4)$ since the non-common edges belong to the same semi-plane.

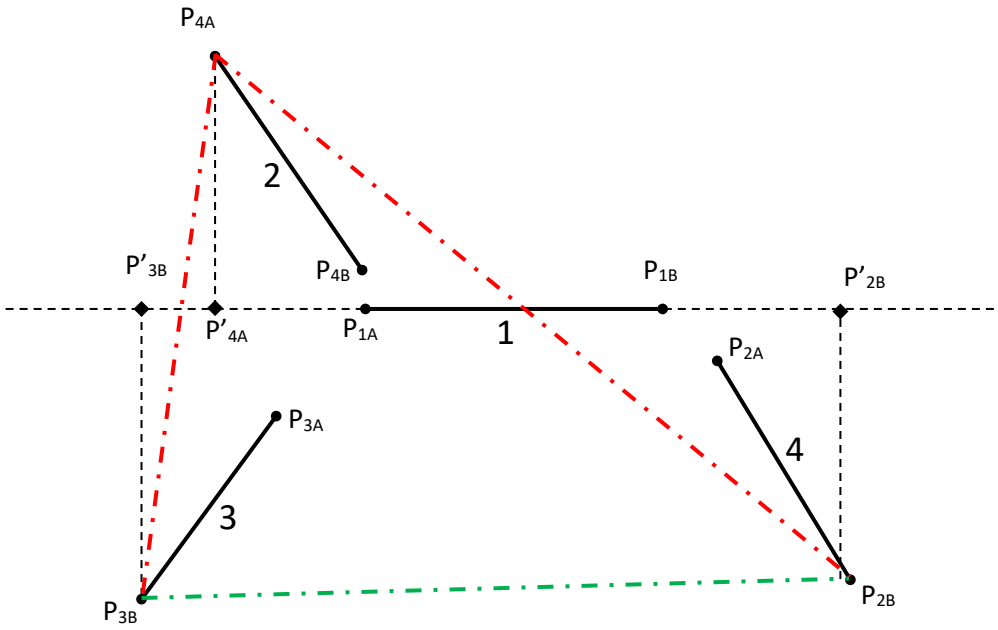


Figure 3.8 Schematic representation of the geometric constraints that define the polygonal triad configuration.

The closed polygonal tetrads are extracted by coupling polygonal triads with two segments in common. An example is shown below.

Consecutively, all the extracted 2D high-level features, grouped by geometric affinity, are stored in a database while, consequentially, all the possible 3D high-level features are stored in a database of model features. A dedicated MATLAB function merges the two sets of features and proceeds to select the detected elements that belong to the most complex and the simplest group and combining them with the respective 3D features. The only exception is represented by parallel triads which provide enough points to ensure a unique solution all by themselves, therefore, they are not coupled with a lower complexity group. The totality of combinations is stored in a match matrix.

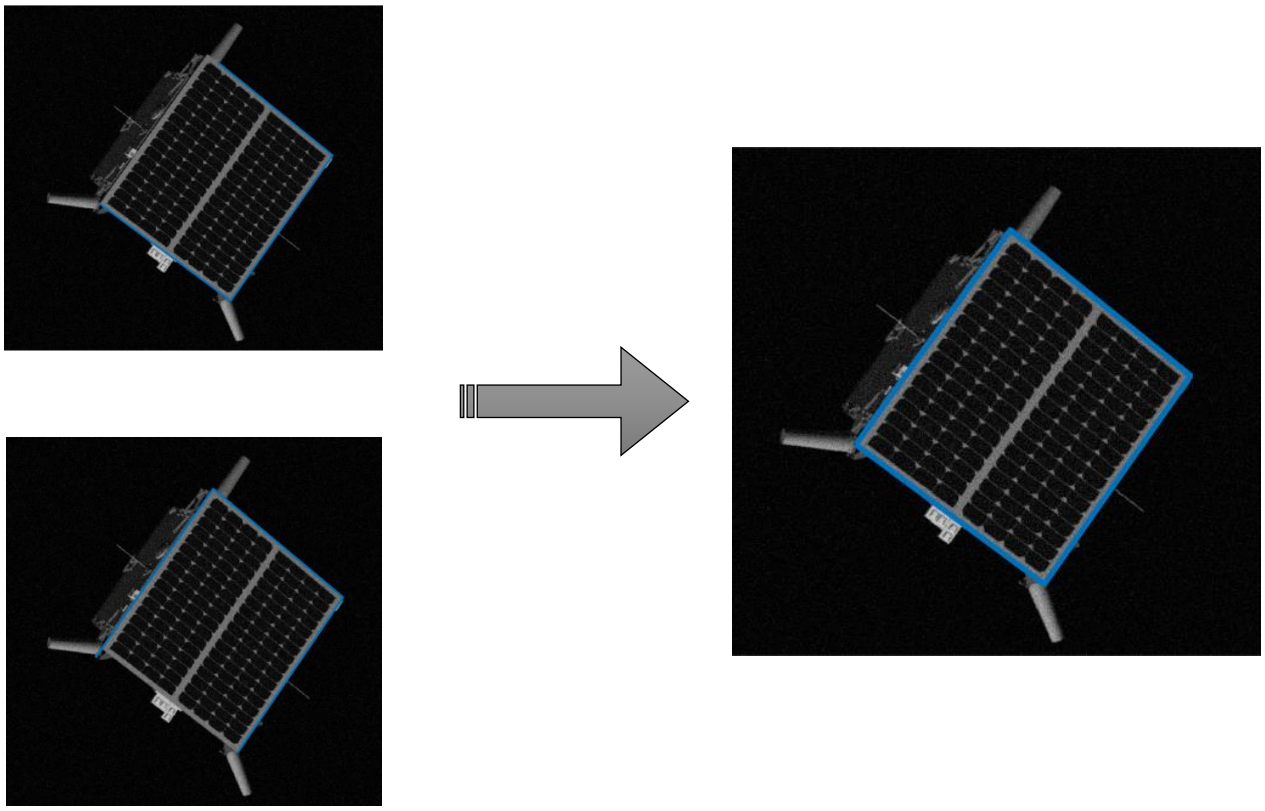


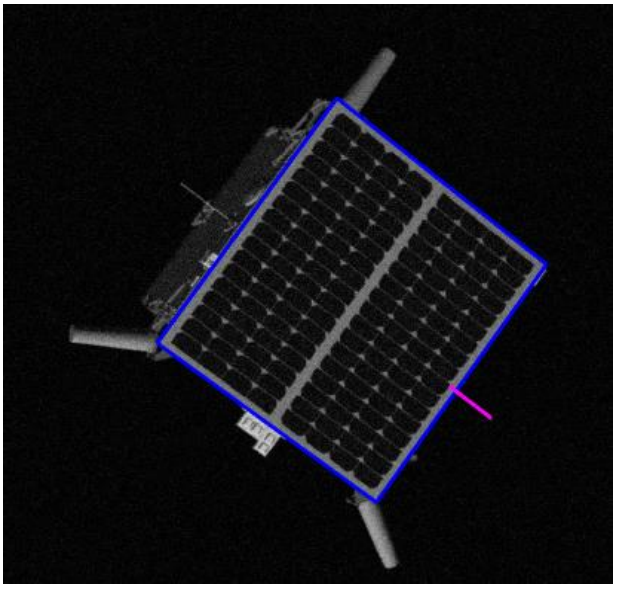
Figure 3.9 *Combination of two polygonal triads into a polygonal tetrad.*

3.1.5 Pose solution and refinement

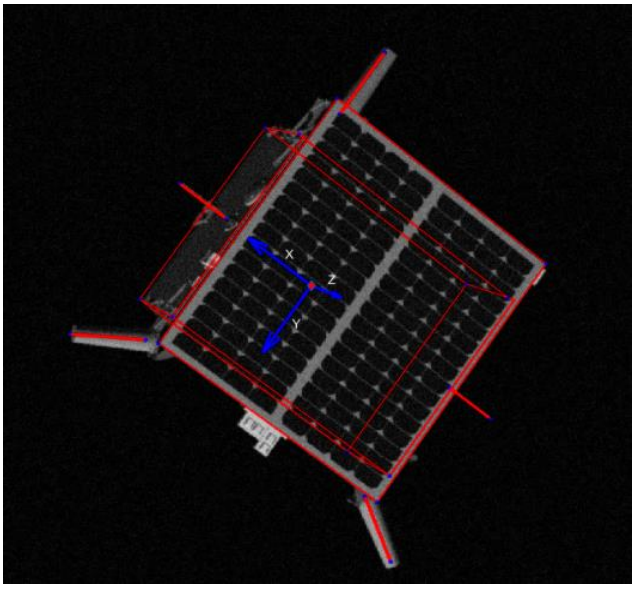
The list of matched features is examined completely to individuate the pose solution with the lowest reprojection error E_{2D}, as in **Eq. 2.13**. Instead of employing the EPnP pose solving algorithm and the Newton-Raphson optimization as suggested in the paper **[31]**, the implementation of the SVD method whose validation is discussed in this thesis relies on a combination of the RANSAC P3P pose solver and non-linear optimization to achieve the solution of the perspective-*n*-point problem. The main disadvantages of the EPnP solver consists in its lack of robustness against outliers, especially in such cases like the ones explored in this thesis where the number of corresponding features is never greater than six and the algorithm incurs frequently in inputs with possibly incorrect feature matching and extraction. On the contrary, the built-in MATLAB function *estimateWorldCameraPose* returns orientation and position of a calibrated camera by means of a P3P solver **[45]** coupled with a RANSAC algorithm for outlier rejection **[53]**. The P3P solver relies on three points to produce up to four symmetrical solution or less if some geometric and algebraic conditions are met, moreover the sample consensus process compares different solutions on the basis of total reprojection error and selects the set of points with the best accuracy. Furthermore, better accuracy is provided by the function *lsqnonlin* which consists in a Levenberg-Marquardt (LM) non-linear optimization, e.g., a

damped least-squares method to find local minima **[54]**. The combinations of high-level features are fed to the pose solver and the five estimates with the best accuracy are subjected to the optimization process and, lastly, the solution with the lowest reprojection error is selected.

In the case of the frame below (**Fig. 3.10**), the total number of feature correspondences to examine is 1920, the procedure that comprises EPnP and Gauss-Newton optimization requires 6.646 s to produce a correct pose solution with a minimum reprojection error equal to 9.505 px, while the P3P+LM pose estimation provides the correct solution, given the same combination of features, a polygonal tetrad and an antenna, with an optimal error of 1.810 px in a total time of 8.389 s.



(a)



(b)

(c)

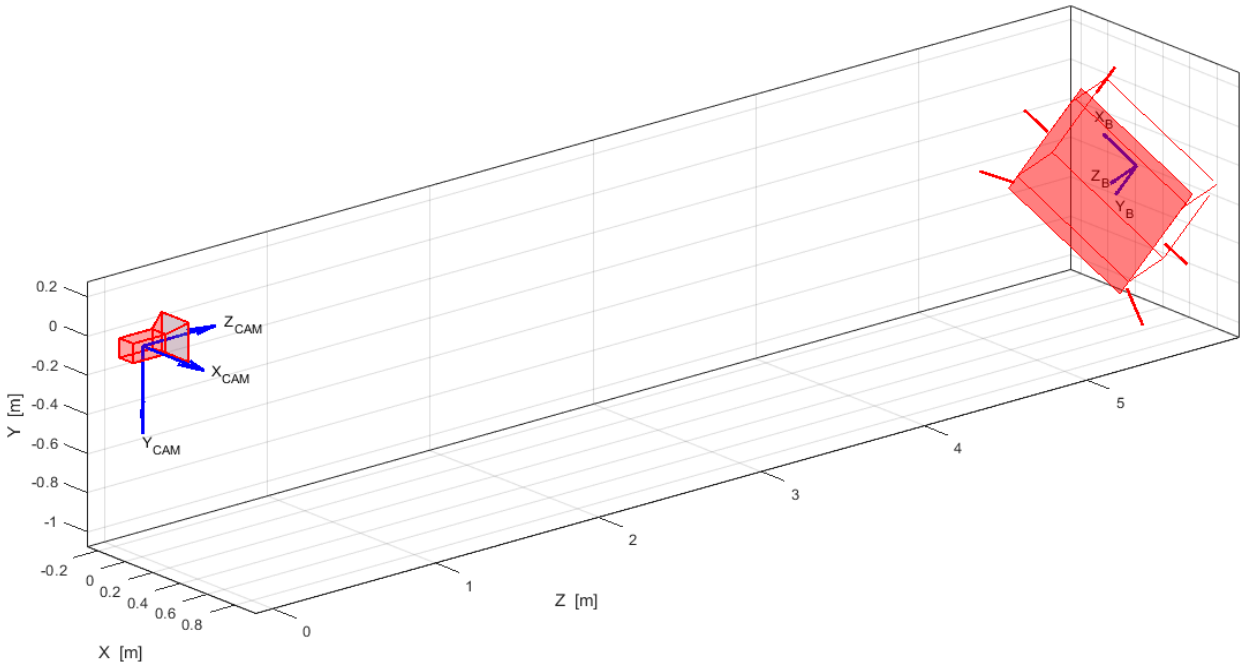


Figure 3.10 An example of correct pose initialization is shown. In (a) the combination of complex and simple features that leads to the best solution is presented, in (b) it's possible to observe how the reprojected wireframe model matches the vehicle structure and, finally, a 3D reconstruction of the relative pose is displayed in (c).

3.2 Validation results

The validation process consists in applying the version of the SVD algorithm previously described to the entirety of the SPEED data. The main purpose is to provide more information regarding the effectiveness of the proposed pose initialization technique and assessing the relation among the accuracy of ROI detection and pose solution processes and the variables of relative distance and orientation.

Given the true attitude R_{true} (rotation matrix) and position vector T_{true} and the estimated R_{est} and T_{est}, the main performance metrics are:

- *Rotational accuracy*: attitude precision is determined by means of the Euler angles representation of the rotation matrix R_{diff} between the true pose and the estimated one. The angles (ϕ, θ, ψ) represent, respectively, the errors about the axes of the camera reference frame $\{X_{CAM}, Y_{CAM}, Z_{CAM}\}$.

$$R_{diff} = R_{est} \cdot (R_{true})^T \tag{3.7}$$

$$R_{diff} \Leftrightarrow (\phi, \theta, \psi) = ERR_R \tag{3.8}$$

  Moreover, the rotation error can be summarized with the following equations:

$$E_R = \cos^{-1}\left(\frac{tr(R_{diff}) - 1}{2}\right) \tag{3.9}$$

- *Translational accuracy:* the accuracy of the relative position calculation is measured as the absolute difference of the vectors T_{est} and T_{true} and summarized by its Euclidean norm.

$$ERR_T = |T_{true} - T_{est}| \tag{3.10}$$

$$E_T = \|ERR_T\| \tag{3.11}$$

- *ROI Intersection over Union* (IoU)*:* the IoU measures the accuracy of area overlapping between the detected ROI and the true ROI which is determined by the limits of the reprojected true model points.

$$IoU = \frac{A_{TP}}{A_{TP} + A_{FP} + A_{FN}} \tag{3.12}$$

The definition of true positive, false negative and false positive areas, respectively A_{TP}, A_{FN} and A_{FP}, is shown in **Fig. 3.11**.

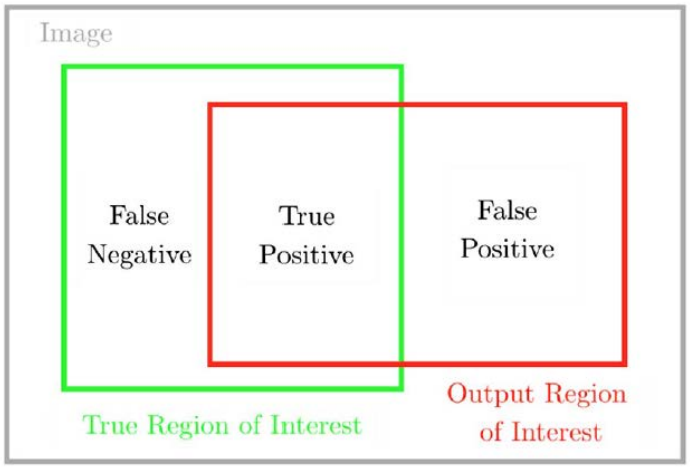


Figure 3.11 Class definitions for the evaluation of ROI detection performance. **[31]**

- *Percentage of pose solutions* (N_{Pose})*:* it represents the proportion of successful pose solution, whether correct or not, over the totality of attempts.
- *Success rate:* the percentage of high-confidence poses in respect to the total number of pose outputs. The classification into high-confidence and low-confidence solution is based on the following thresholds:

$$\|E_T\|_2 < 30\ cm \qquad\qquad \|E_R\|_2 < 10° \tag{3.13}$$

- *Number of line correspondences* (N_{lines})*:* the number of detected segments that match correctly with the true edges of the spacecraft figure. A correct

- correspondence is found when the distance between detected and true endpoints is smaller than $0.025 \cdot l_{ROI}$.
- *Number of point correspondences* (N_{points})*:* the number of correctly detected endpoints. With respect to N_{lines} , the value of point correspondences considers fragmented and partially detected segments.
- *Number of combinations* (N_{comb})*:* number of possible matches between 2D and 3D high-level features that are tested for pose determination.
- *Computational Time* (t_{comp})*:* computational time required to obtain a pose estimation

The mean results for the SPEED images are summarized in **Tab. 3.2**. The results contained in **[31]** show that over a simple dataset of 25 images the SVD method produced an high-confidence solution only for 5 of them and a pose solution for 13 of them (**Tab. 2.1**). The performance of the algorithm is similar to the one exposed in the paper in terms of percentage of pose solutions, translational and rotational accuracy and computational time, yet there is a clear difference in success rate. Unfortunately, the dataset employed in **[31]** it's not sufficiently large and diverse to provide a reliable evaluation of the algorithm implemented in the thesis, in fact, only four images present a composite background and Tango's attitude and position are consistently similar throughout the images. Let it be noted that in accordance with the typical limitations of vision-based measurements the accuracy along and around the boresight of the camera $\{Z_{CAM}\}$ is greater than for other axes.

In order to provide a better analysis of the algorithm's performances two subset of the SPEED images will be considered: the first half of the dataset (images from 1 to 6000) whose images present an empty background will be referred to as SPEED-1 and the second half which comprises of images with a Earth surface background will be referred to as SPEED-2.

Success Rate	7.99 %
$\boldsymbol{N_{Pose}}$	59.42 %
Mean IoU	49.50 %
Mean E_T [mm]	93.75
Mean E_R [deg]	2.85

Mean ERR_T [mm]	[9.42 9.66 90.79]
Mean ERR_R [deg]	[0.77 1.65 1.69]
Median N_{lines}	1
Median N_{points}	3
Mean N_{comb}	2875
Mean t_{comp} [s]	9.6615

Table 3.2 SVD architecture performance metrics. The accuracy values refer to high-confidence solutions only, while the remaining parameters are calculated over the entire dataset.

The results collected in **Tab. 3.3** demonstrate how a composite background can greatly affect the effectiveness the functionality of the Weak Gradient Elimination technique for ROI detection. The presence of Earth's surface in the background produces spurious elements in the gradient binary image and if the filtering process fails to eliminate them the limits that define the satellite's bounding box are incorrectly located. An erroneous identification of the target's ROI translates into the implementation of biased geometric hyperparameters for the Hough transform, thus leading to a wrong set of detected 2D linear features and a reduced number of point and line correspondences. Moreover, the consistent difference between the values of N_{Pose} and success rate for the two dataset SPEED-1 and SPEED-2 proves the dependency of the SVD method's effectiveness on the correct perceptual feature organization, which relies on ROI's dimensions. Nevertheless, the accuracy of pose outputs, both translational and rotational, it's not affected by the Earth's background interference when the conditions for high-confidence are met.

Some plots are presented hereinafter to further analyse the variations of performance metrics in regard to the main variables of the pose initialization problem, such as relative distance, bearing angles, feature combinations and line correspondences.

	SPEED-1 dataset	SPEED-2 dataset
Success Rate	8.85 %	4.96 %
N_{Pose}	92.62 %	26.22 %
Mean IoU	82.63 %	16.37 %
Mean E_T [mm]	90.69	113.05
Mean E_R [deg]	2.94	2.31
Mean ERR_T [mm]	[9.36 9.66 87.70]	[9.75 8.93 110.28]
Mean ERR_R [deg]	[0.78 1.68 1.78]	[0.69 1.49 1.14]
Median N_{lines}	2	1
Median N_{points}	5	2

Table 3.3 SVD architecture performance results are summarized in this table. The accuracy values refer only to high-confidence solutions, while the remaining parameters are calculated over the entire dataset.

Relative distance is a fundamental variable of pose estimation and, in fact, the quality of an algorithm is deemed optimal if its accuracy and robustness are considerably predictable over a certain range of operativity. In the case of the SVD architecture the variation of E_T and E_R along with the increase of range follows a typical pattern for vision-based algorithm. As it's displayed in **Fig. 3.12**, both the translational and the rotational estimation accuracy tend to diminish gradually over the range from 3 m to 15 m, while the results for longer ranges don't fit a specific trend given the fact that the low number of correct pose solution doesn't have statistical merit. Together with the error, ROI's IoU metric and pose solver's success rate decrease according with the increase of relative distance between 3 m and 25 m, as shown in **Fig. 3.13**.

The relation between the elevation and azimuth angles and the outputs of ROI detection and pose estimation is shown in **Fig. 3.14**. High-confidence solutions and correct ROIs are distributed in accordance with the built-in randomly generated distribution of SPEED data without any substantial predominance in elevation or azimuth.

In addition to the previous elements, in **Fig. 3.15 - 3.16** the results of the SVD architecture are evaluated in relation to the various combinations of high-level

features. In the first plot the success rate of different combinations is summarized in an histogram. Along the horizontal axis the pairs of features are ordered by the number of high-confidence solutions provided, from the most common (polygonal tetrad and antenna) to the least frequent (proximal pair and antenna). As it can be observed the combinations of an antenna with more complex groups such as polygonal tetrads and triads and parallel pairs provide commonly more correct pose outputs, thus, demonstrating the main hypothesis that combining an easily detectable, characteristic, simple feature with a complex, rarer element of the target's structure confers sufficient robustness to pose estimation algorithms while reducing the search space. Besides, translational and rotational accuracy proves to be essentially uninfluenced by the type of feature groups employed by the pose solver subsystem.

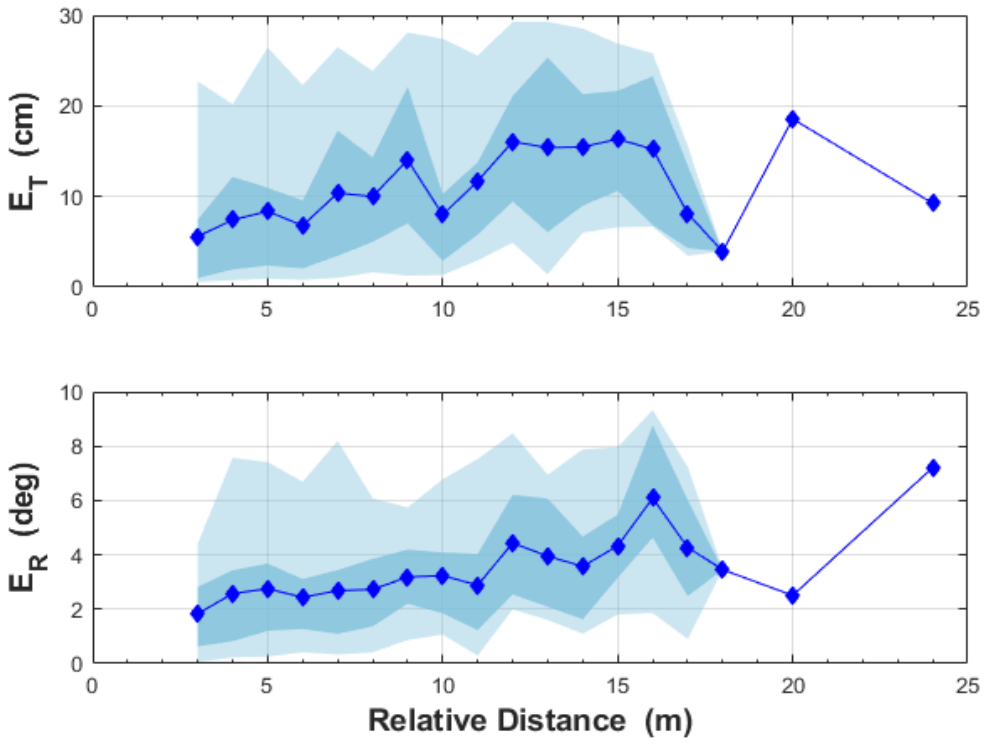


Figure 3.12 *Pose estimation translational and rotational errors in relation to relative intersatellite range. The shaded regions show the 5, 25, 75 and 95 percentile values.*

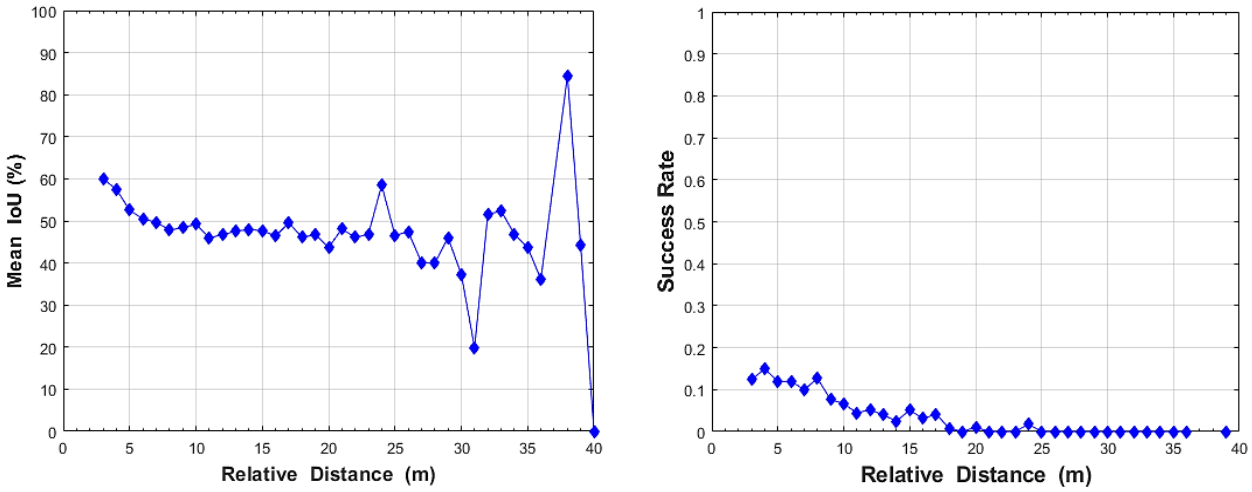


Figure 3.13 *Mean IoU and success rate values in relation to relative intersatellite range.*

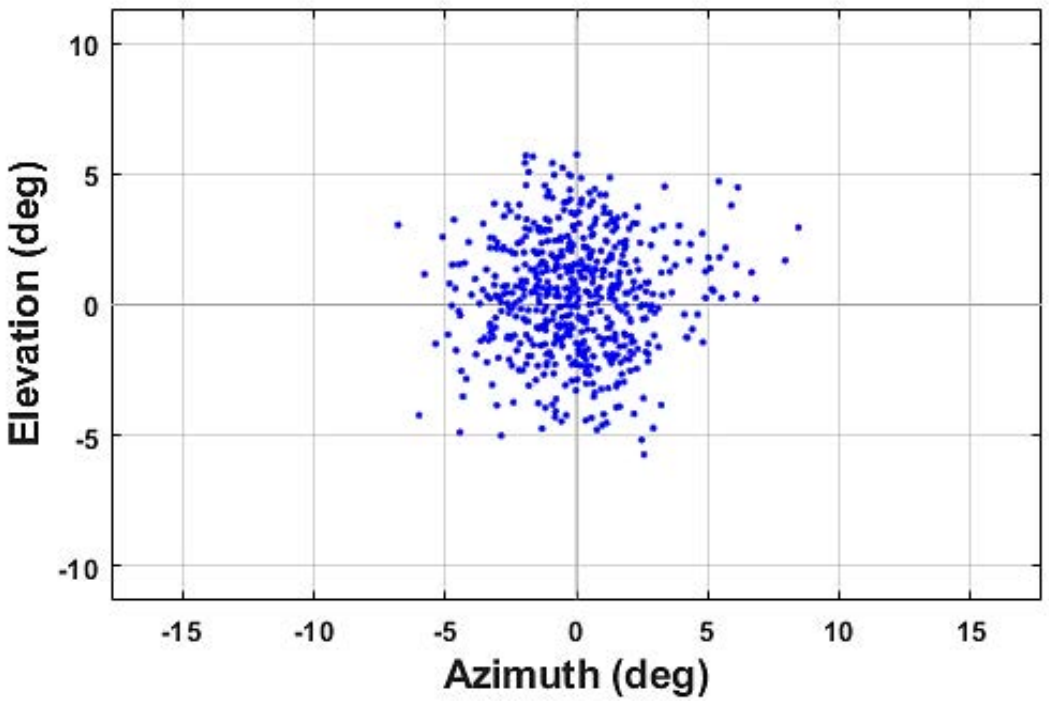


Figure 3.14 *Distribution of high-confidence pose solutions ROI centers across the Field of View.*

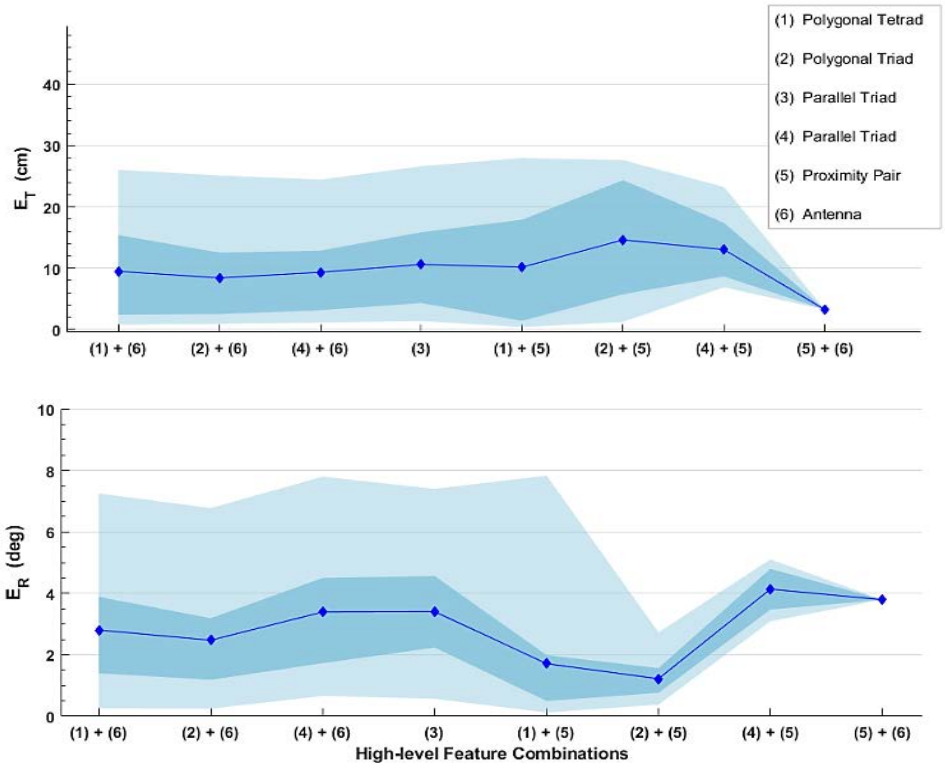


Figure 3.16 *Translational and rotational errors of high-confidence pose solutions divided by the combinations of geometric features. The shaded regions show the 5, 25, 75 and 95 percentile values.*

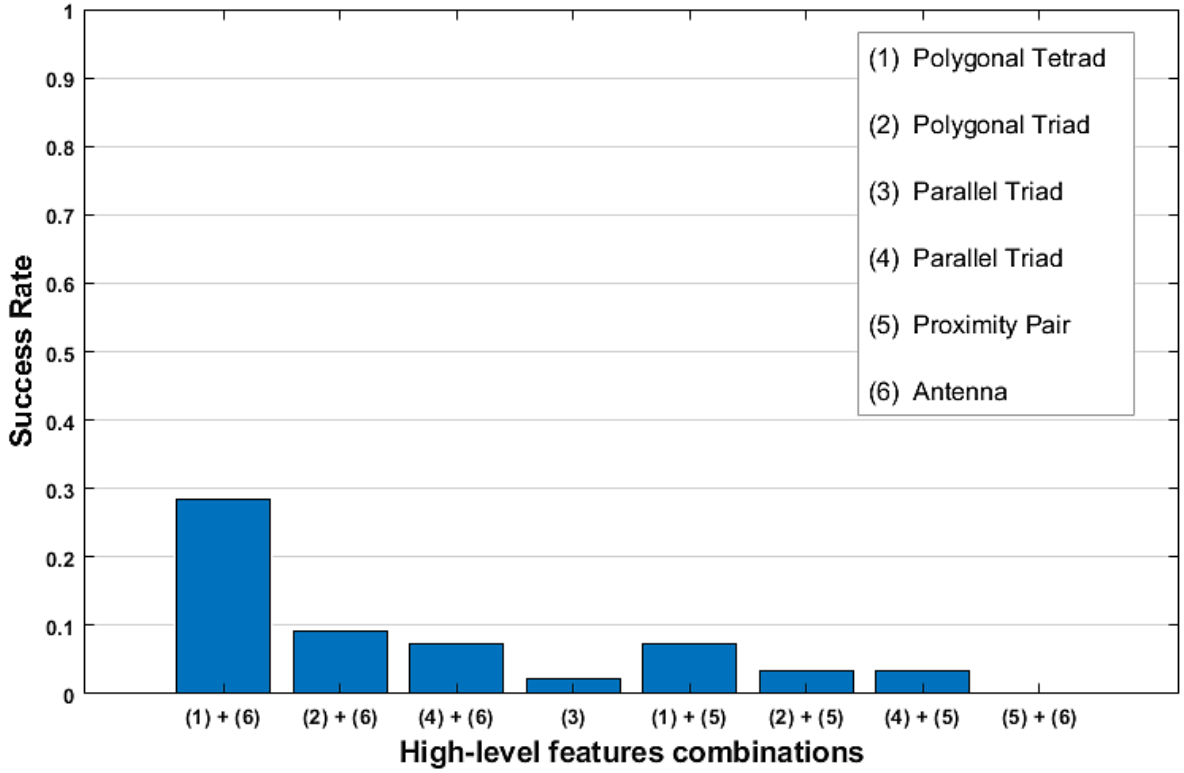


Figure 3.15 *Pose estimates are divided by feature groups and the respective success rate is plotted against the most common combinations.*

In order to completely estimate the validity of the SVD method, a series of plots that correlate the efficiency of single subsystems to the final results are presented.

Fig. 3.17 shows the relation between the precision of the ROI detection process and the success rate of the pose solver. The identification of a correct bounding box for the target's shape within the image leads generally to an increased probability of producing a high-confidence estimate, since the edge detection process, as well as the feature synthesis block, rely on geometric thresholds derived directly from the dimensions of the ROI. **Fig. 3.18** shows the precision of the estimated coarse range in relation to the true relative distance T_{true}, as calculated after the ROI identification. ROI based measurements of range and LOS are sufficiently accurate to provide an initial guess for pose refinement and fast, reliable data for medium-range proximity navigation.

Furthermore, **Fig. 3.19** displays the variance of the translational and rotational errors in regard to the number of line correspondences, while in **Fig. 3.20** the success rate distribution is plotted against the same variable. The number of true edges (N_{lines}) detected by the feature extracting block of the algorithm is a metric of its effectiveness which is directly related to the results of the pose solver technique. In the case of the

SVD architecture, though, the input to the PnP problem it's always composed only by two geometric groups which amount to a number of five 2D segments at best. Hence the performances of the pose initialization in terms of accuracy and success rate increase along with the number of line correspondences until a maximum of five lines to remain fundamentally unchanged, thereafter, for $N_{lines} > 5$. The same conclusions can be drawn as regards the relation between point correspondences and success rate for $N_{points} > 10$.

Lastly, **Fig. 3.21** shows a set of high-confidence and low-confidence solutions and ROI-only estimations for some of the images belonging to the SPEED set. The pose initialization technique proves to be successful in a variety of cases with low-light conditions, different ranges and attitudes and in the presence of a planetary background, yet it lacks sufficient robustness to provide consistently similar results in the same conditions. For instance, the images 11483, 9417 and 11965 present a composite background due to the presence of Earth's curvature horizon which affects the proper detection of the satellite's region of interest and introduces the identification of spurious edges, yet the SVD architecture produces a correct pose solution only in one of them, failing to assess the presence of high-level features in the other two. Likewise, both the images 6722 and 9771 show the target in low lighting conditions and, despite a better ROI detection, the pose initialization doesn't provide a correct estimate to the first one, while it calculates the pose within the confidence error for the second one.

Additionally, the graphs in **Fig. 3.22** show the spatial distribution of the high-confidence pose across the horizontal and vertical planes of the target reference frame. The plots demonstrate the absence of preferential directions for pose determination around the axis $\{Z_{body}\}$ and, instead, show that the high-confidence solutions are statistically more frequent if the target's larger surface, either the solar panel or the bottom surface of the main body, it's captured by the camera. Given the configuration of the Tango spacecraft, which is composed by two easily identifiable polygonal shapes and a series of five antenna asymmetrically arranged parallelly to the X-Y plane, and considering that the combination of a polygonal tetrad with an antenna offers usually a higher success rate, as shown in **Fig. 3.15**, it's reasonable that the majority of high-confidence results doesn't lie on the horizontal plane of the spacecraft but it's distributed along the $\{Z_{body}\}$ axis in both directions. For instance, a configuration like the one displayed in **Fig. 3.10** is optimal since it allows for an easier identification of the antennas and the solar panel at the same time.

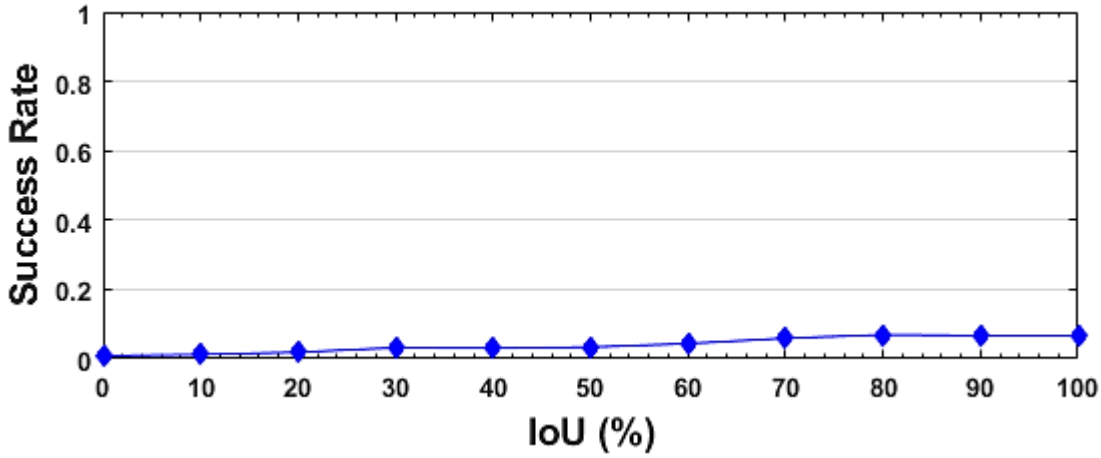


Figure 3.17 *The plot portrays the success rate variation over the percentage of ROI overlapping. The ROI detection process, when effective, enables better overall performances for the pose initialization algorithm.*

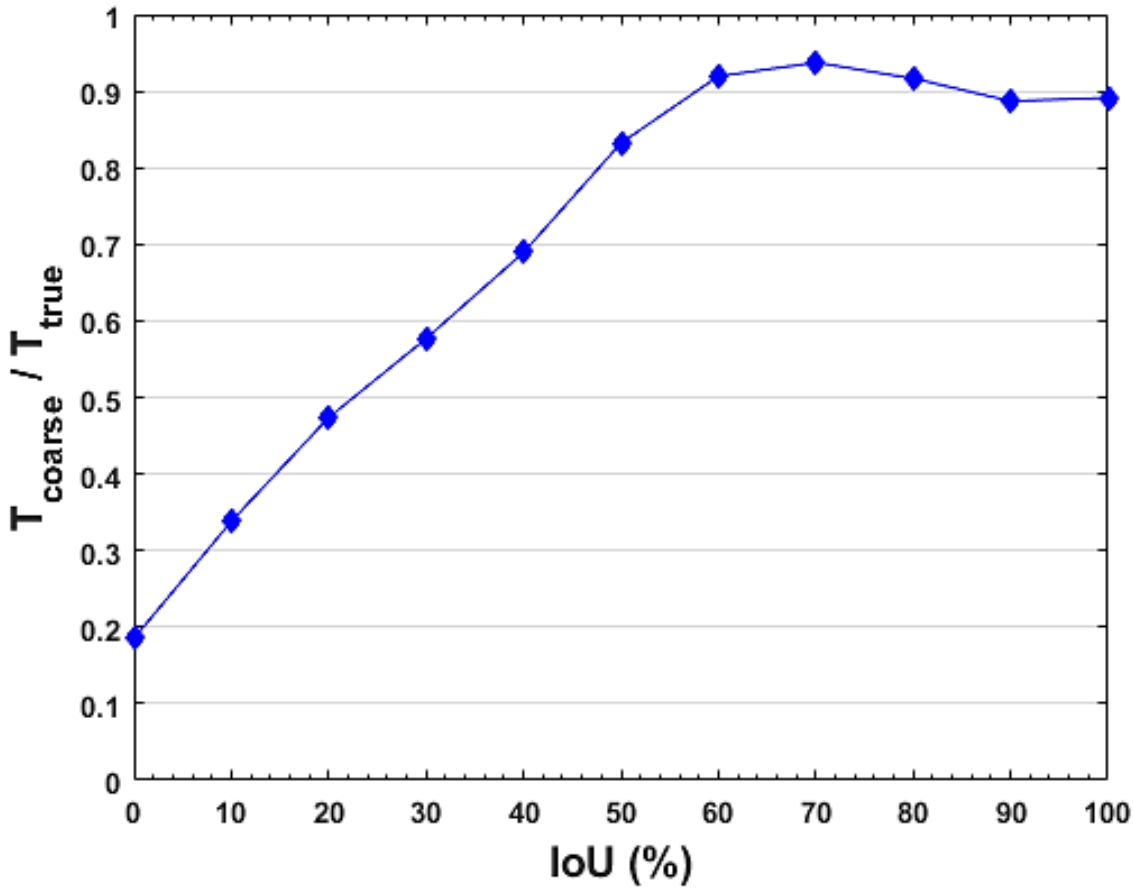


Figure 3.18 *The precision of the range coarse estimation is related with IoU of the ROI, as shown in the plot. The ROI-based measurements are already sufficiently accurate for a 60% area overlapping.*

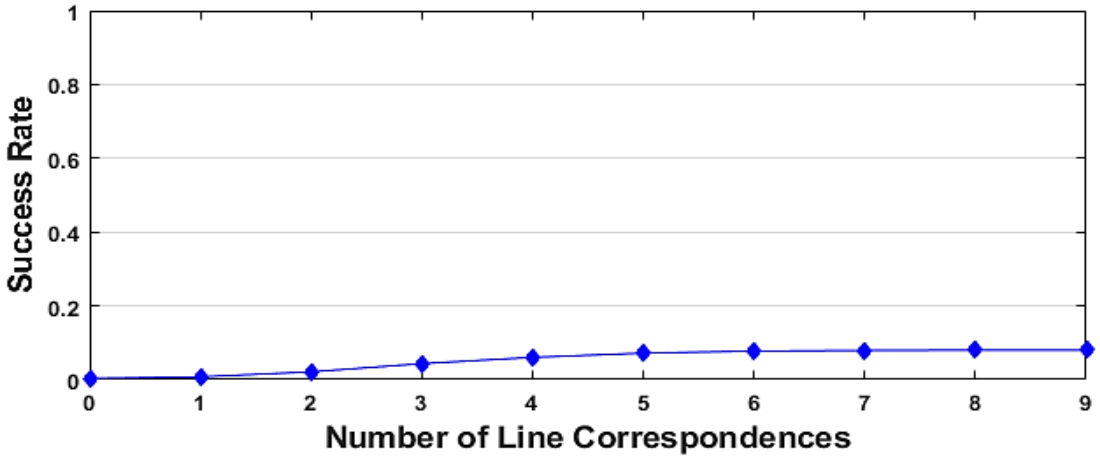


Figure 3.19 *Success rate plotted against the mean number of correspondences between true edges and detected segments.*

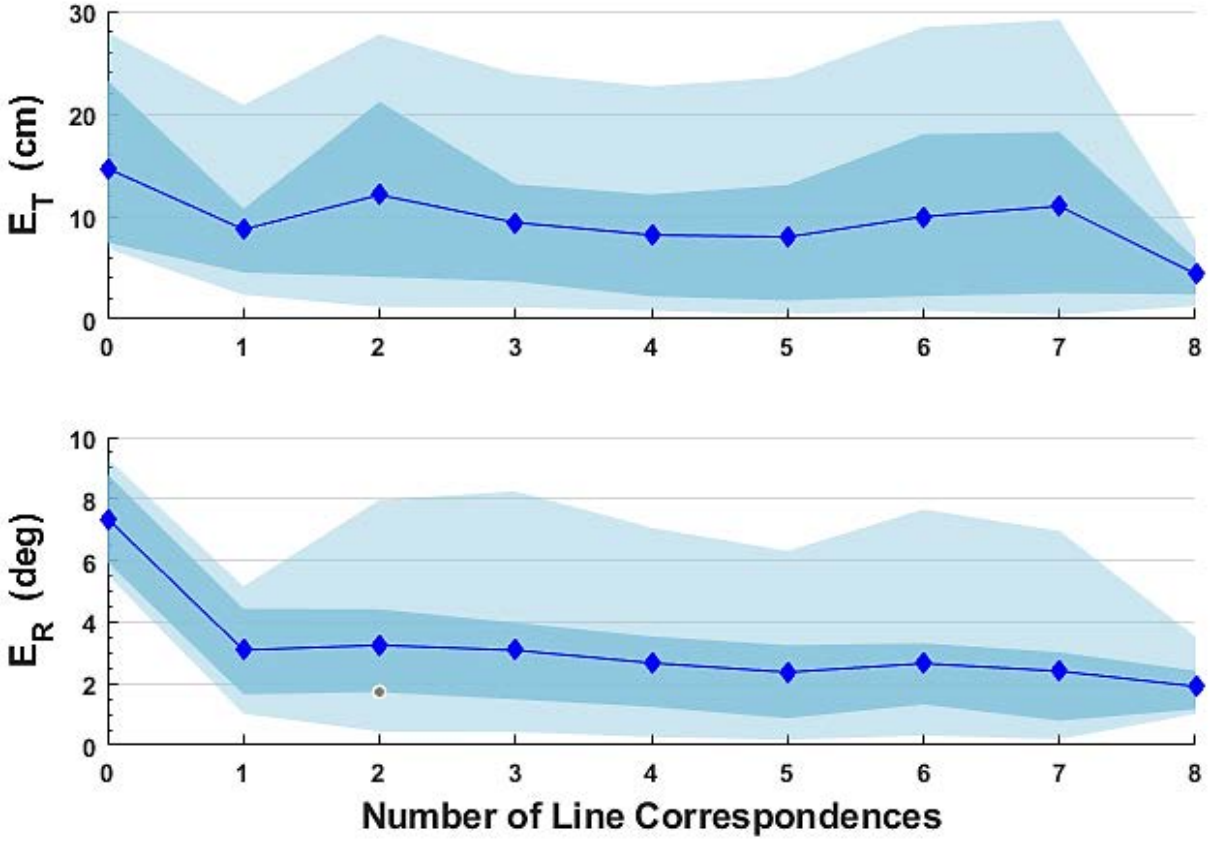
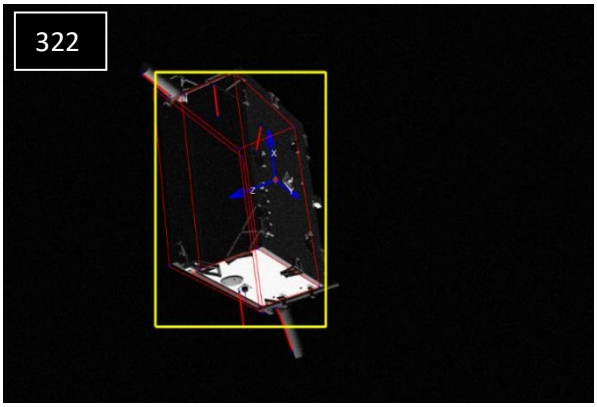
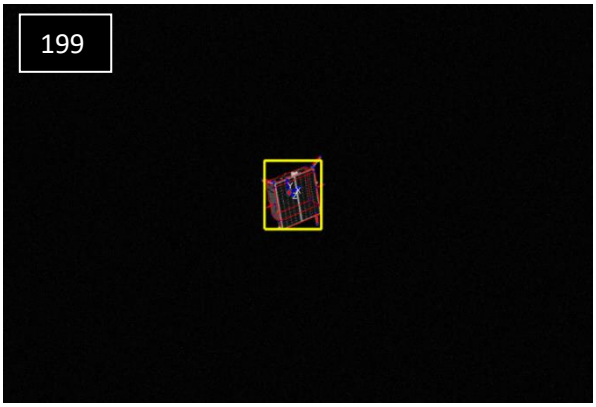
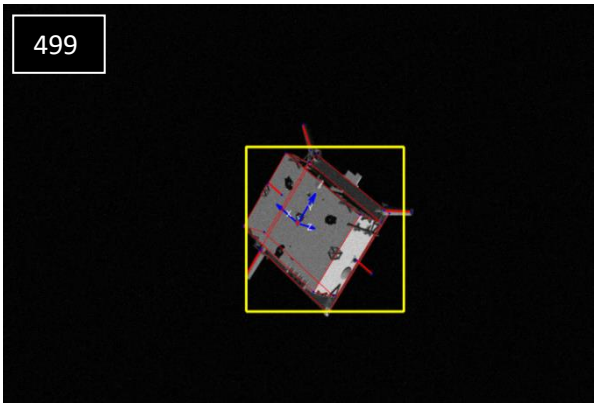
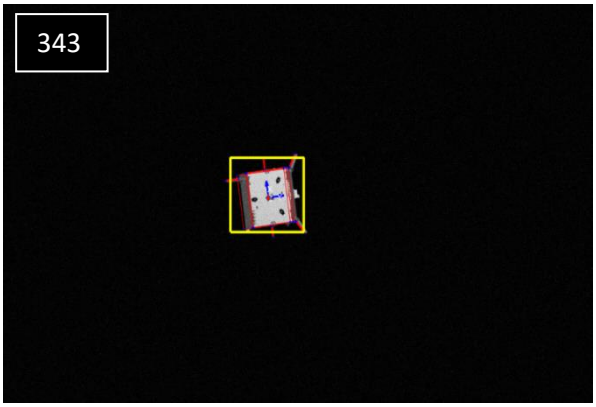


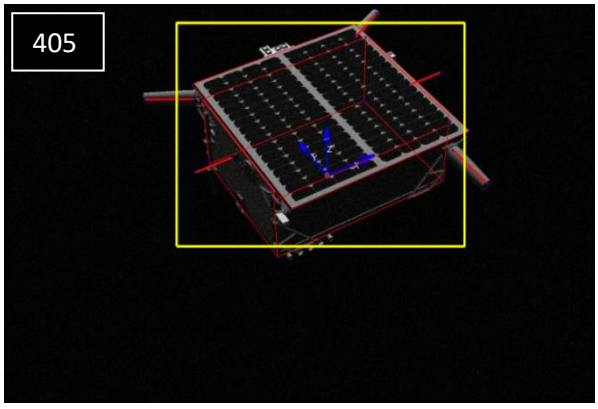
Figure 3.20 *Translational and rotational errors of high-confidence pose solutions sorted by the number of line correspondences. The shaded regions show the 5, 25, 75 and 95 percentile values.*

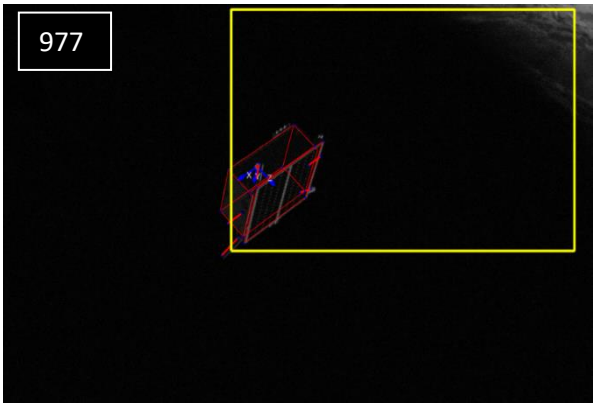

322

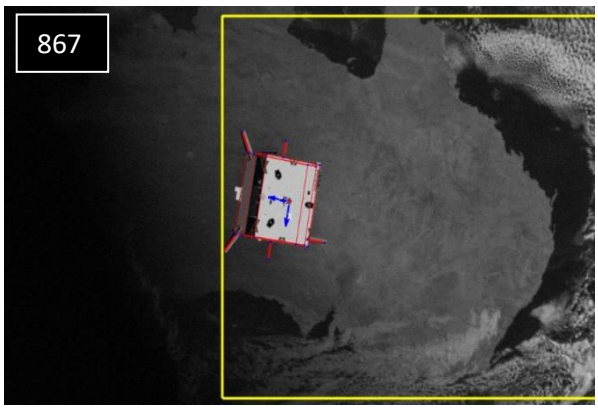

199

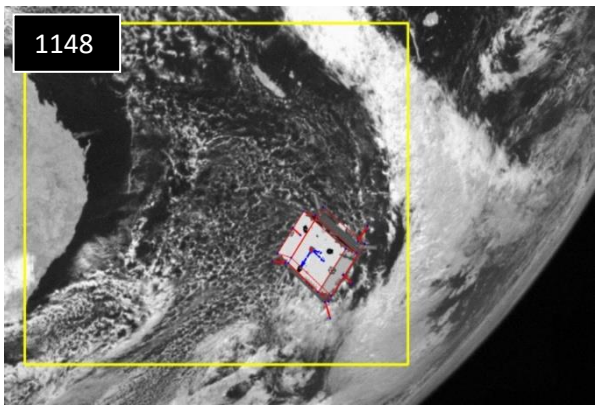

499


343


405


977


867


1148

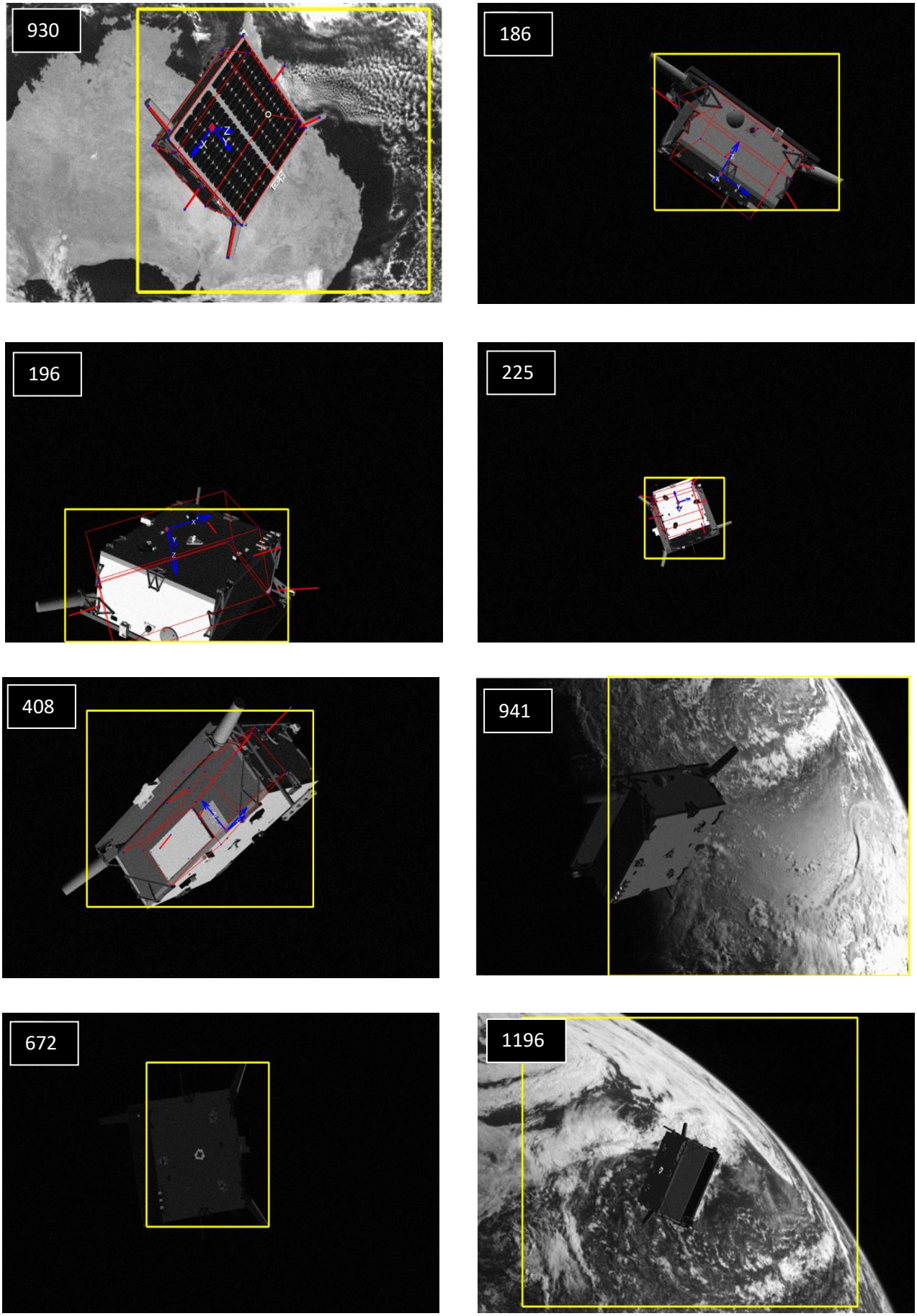
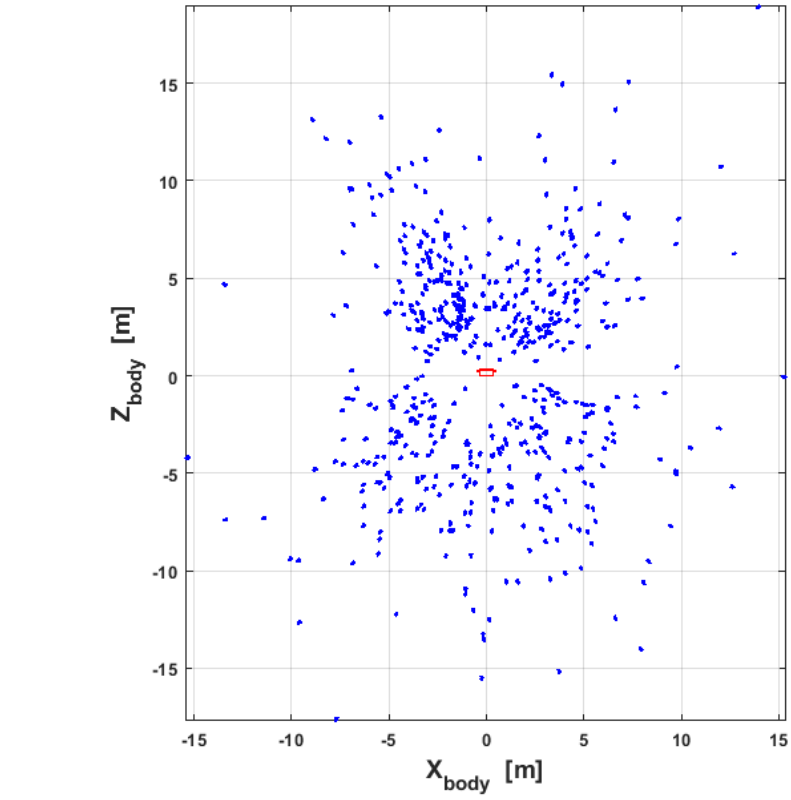
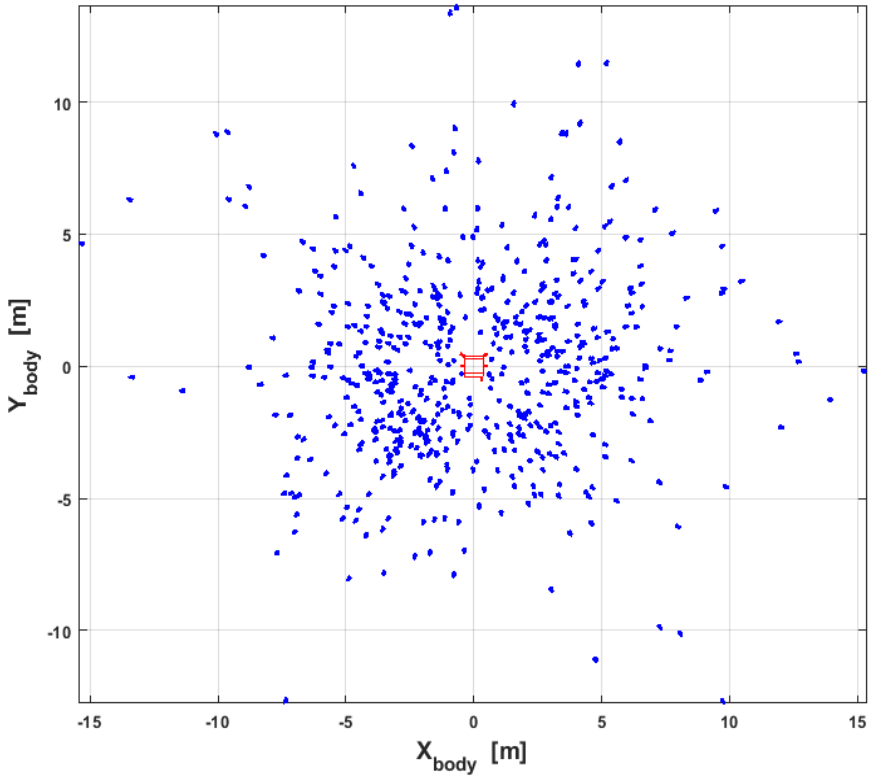


Figure 3.21 *Pose initialization results using the SVD algorithm.*



(a)



(b)

Figure 3.22 *Spatial distribution of high-confidence pose results.*

Chapter 4

Algorithm validation on experimental imagery

In this chapter an experimental setup for relative pose estimation is presented. After implementing and testing the SVD architecture on the SPEED images, further validation tests and analysis are conducted to assess the robustness of this method in more restricted conditions with a real experimental setup. The process of creating a database of images for a CubeSat mock-up in the SPARTANS facility is described. Lastly, an analysis of the algorithm's performance is presented.

4.1 Experimental setup

A second series of validation tests are conducted with a mini-satellite mock-up, a dataset of images captured by a stereo-vision system and a test-bed for satellite relative motion detection. The experimental setup is instrumental to define a dataset of images and reference attitude and position parameters to measure the performance of SVD pose initialization.

The experimental facility employed to create the dataset is SPARTANS, a testbed for satellite relative dynamics, under continuous development by the Center of Studies and Activities for Space "G. Colombo" (CISAS) at the University of Padova and functional since 2010. The facility has been created to test the accuracy of pose estimation techniques and demonstrate the effectiveness of Guidance and Navigation Control (GNC) architectures. SPARTANS comprises of:

- Glass-covered flat surface (2 x 3 m)
- External Control Station (computer)
- Two or more Spacecraft Simulators

- Global Navigation System

The facility employs six IR cameras and IR reflective spherical markers, along with a dedicated software, to provide a global reference navigation system and assess the orientation and translation parameters of the Spacecraft Simulators. The simulators consist in an Attitude Module (AM) that provides rotational control by means of mechanical gimbals and a Translational Module (TM) that operates along the 2D axis of the low-friction glass table by means of an air-cushion system.

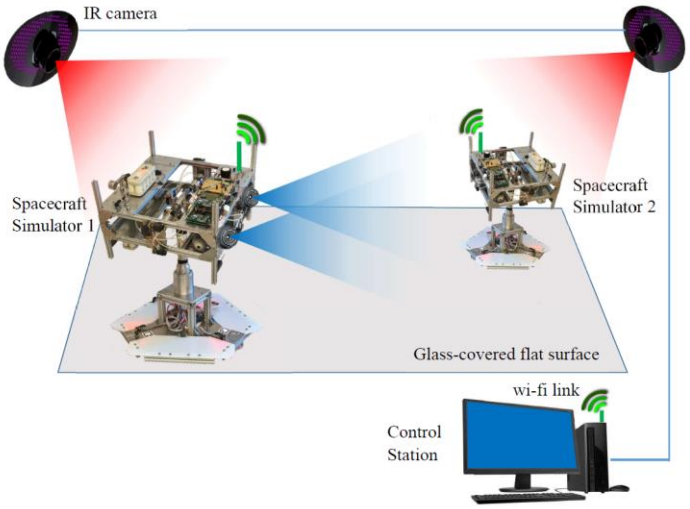


Figure 4.1 *Overview of the SPARTANS testing facility.* **[55]**

In order to measure the accuracy of the SVD technique only the Global Navigation System has been employed, so to compute the "true" attitude and position values of both the satellite prototype and the camera frame by means of a motion capture algorithm. As demonstrated in **[55]**, the system computes the degree of freedom (DoF) of the simulator with a precision inferior to 0.1° and 1 mm, respectively for attitude angles and single-point spatial coordinates, therefore it offers a sufficiently good reference estimate of the true pose.

The satellite mock-up consists in a simplified 1:1 model of a 2U-CubeSat, a common format for miniaturized satellites widely used in cost-effective LEO missions. The model simulates typical features of CubeSat satellites, such as deployable and body mounted solar panels, simple and modular structures and long UHF antennas. Since the intent of the CubeSat mock-up is to simulate the appearance of a real vehicle but not the functionality, the simplest choice for solar panel is the adoption of printed pattern glued to plexiglass plates while in regard to other uncovered surfaces and adhesive reflective layer has been used. The prototype is shown in **Fig. 4.2**.

Figure 4.2 *Back and front face of the CubeSat model with the main solar panel, two antennas and three IR reflective markers on the top surface.*

The model is mounted on a aluminium support that is fixed to a STANDA 8MR190-90 motorized rotative stage (**Fig. 4.3(a)**). The rotor is activated by a stepper motor which moves a transmission worm gear. The actuator allows to impose a finely controlled motion to the CubeSat model and simulate rotation along the main axis of the body frame. In order to avoid interference with the target in the foreground, the background of the scene is covered with a black low-reflective cloth and the model support frame is black painted.

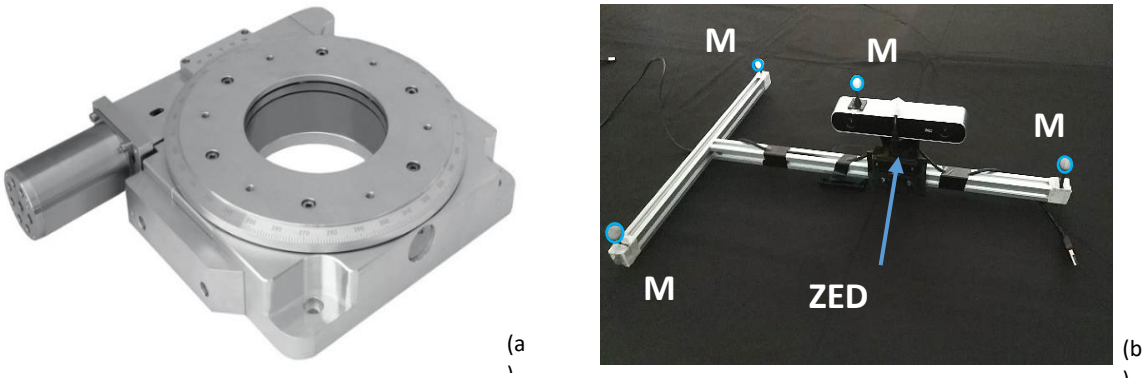


Figure 4.3 *The rotative stage (a) and the ZED camera mounted on its own support frame with four reflective markers M_0 , M_1 , M_2 , M_3 (b).*

Six reflective markers are attached to the assemble of the mock-up and the support plate, three at the top of the model and other three on the base. Assuming that the mock-up is rigidly fixed to the support frame, the relative position of the six markers remains constant (**Fig. 4.6**).

The camera setup consists in a ZED stereo-camera system rigidly mounted on a aluminium frame attached to an adjustable tripod. The stereo-camera main properties are summarized in **Tab. 4.1**. The CubeSat imagery is retrieved by the left camera of ZED's sensor. Four reflective markers are positioned on the camera frame and the camera shell and they are employed to determine the attitude and position of the sensor in respect to the world coordinates (**Fig. 4.3(b)**).

ZED stereo-camera Specifications	
Image Format (full)	$4416 \times 1242\ px$
FoV	110°
Pixel Size	2 µm
Sensor Size	$1/3''$
Shutter	Electronic Synchronized Rolling Shutter
Baseline	120 mm

Table 4.1 ZED stereo-camera principal parameters.

4.2 Calibration process

The motion capture system requires an initialization procedure to define a Global Reference Frame (GRF) and an operative space where the marker's detection is accomplished. First of all, the six IR cameras are directed towards the center of the scene since they are equipped with IR illuminators and the uncertainty of markers coordinates is reduced in relation to the number of cameras that detect them. The GRF is a local vertical-local horizontal frame defined by a set of physical axes, i.e., three sticks oriented in perpendicular directions with markers located at different lengths in order to identify them. The global frame calibrator is displayed in **Fig. 4.4**. Once the process of calibration and optimization is automatically implemented, a so-called World Frame (WF) is determined.

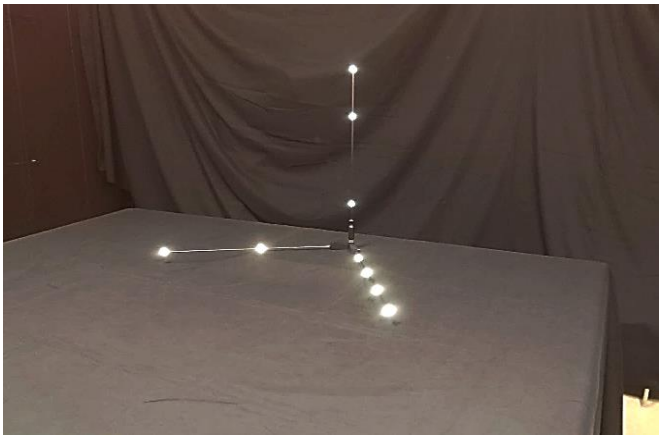


Figure 4.4 *The calibrating physical frame for the Global Reference System.*

Referring to the camera and model markers the minimum number to define a unique reference frame is three. Given the 3D coordinates of three markers M_0,M_1, and M_2, two unit vectors are calculated as follows

$$\hat{p} = \widehat{M_0M_1} \qquad \hat{q} = \widehat{M_1M_2} \tag{4.1}$$

then the linearly independent axes are defined as

$$\hat{X} = \hat{p} \qquad \hat{Y} = \hat{Z} \times \hat{X} \qquad \hat{Z} = \frac{\hat{p}\times\hat{q}}{\|\hat{p}\times\hat{q}\|} \tag{4.2}$$

And the relative orientation and position is defined as follows:

$$T = [M_0] \qquad R = \left[\hat{X}\ \hat{Y}\ \hat{Z}\right] \tag{4.3}$$

Another preliminary procedure must be conducted to calibrate the vision sensor. The ZED camera intrinsic parameters are obtained by using the Camera Calibration Toolbox **[56]** provided within the MATLAB software. The calibration algorithm requires only a series of checkboard images sparsely distributed across the field of view and partially oriented, captured by a stationary point of view. The checkboard consists of a set of 5x8 consecutive black and white squares with a width equal to 110.5 mm. A total of 43 images have been employed to determine the distortion

parameters and the intrinsic matrix of the ZED's left camera. The calibrator software identifies automatically the checkboard's corners (28) in each image and operates a least square optimization over the total reprojection error to compute the camera's intrinsic parameters. A threshold for the reprojection error can be set so to avoid images with blurring that generally bias the optimization results. The calibration procedure provides the results in **Tab. 4.2**:

ZED left camera Parameters	
Image Format	1920 × 1080 px
Focal Length	[1396.0 1397.2] px
Principal point	[1075.2 536.8] px
Radial distortion	[-0.1462 -0.0661 0.0956]
Tangential distortion	[$-5.0148 \cdot 10^{-4}$ $-2.5563 \cdot 10^{-4}$]
Skew	0
Mean reprojection error	0.1741 px

Table 4.2 ZED left camera calibration results: intrinsic parameters and distortion coefficients.

The calibration of ZED's left camera and of the global reference frame for the motion capture system allows to calculate the transformation matrix between the camera support frame, which can be tracked at any moment with the markers detection, and the camera frame, as defined according to the pinhole model. The transformation matrix which describes uniquely a rotation and translation operation between two frames can be calculated by minimizing the reprojection error of a set of well-defined points. By retrieving the 3D coordinates of a set of points and the respective 2D coordinates on the image plane it's possible to determine a transformation matrix. In particular, since the motion capture detection system provides the 3D coordinates of the markers in relation to the World Frame, a Support Frame (SF) can be defined with the calculations shown in **Eq. 4.1-4.2-4.3** among three of the markers attached to the camera support. A number of markers, placed at the centre of the scene, are translated across the field of view of the camera, which is in a fixed position, and multiple images are captured while the Global Navigation System of SPARTANS tracks the

position of the calibration markers and the camera support's markers (**Fig. 4.5**). The 3D coordinates and a set of 30 images are obtained and processed by a dedicated calibration function.

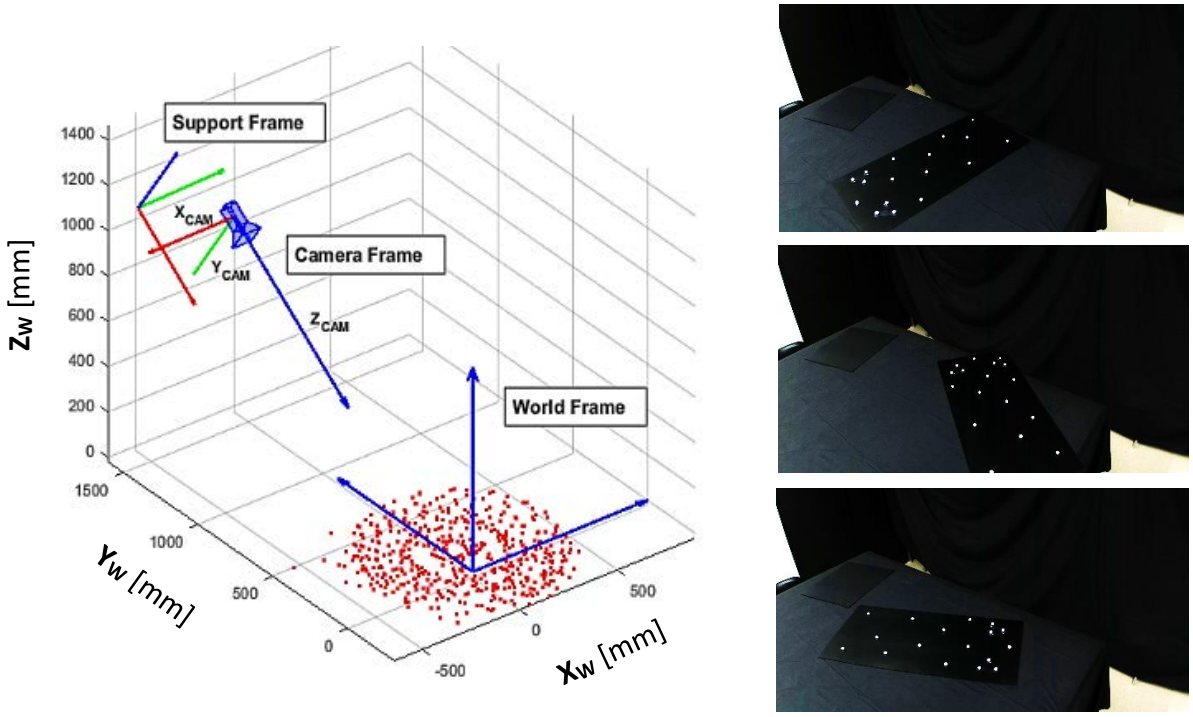


Figure 4.5 *The results for the calibration of the transformation matrix between WF and CF. On the left the configuration of the Camera Frame, the Support Frame and the World Frame with the calibration markers point in red. On the right a montage of the calibration images showing the distribution of markers across the field of view of the camera.*

The algorithm is based on the paper **[59]** and consists in three fundamental sections: point detection and matching and transformation estimation and a final optimization. The main challenge of the calibration process is detecting the markers in the images and correctly matching them with the corresponding 3D points. The detection system relies on a manual association for initialization and subsequently operates by applying a nearest neighbour search based on Euclidean distance between the projected image points, derived by a first guess transformation matrix, and a number of circular features previously detected on the image with the built-in MATLAB function *imfindcircles*. Given the spherical shape of the markers, their appearance and their common geometrical properties their centers are easily identifiable by setting a closed range of radial distribution for the circle detection technique. Hence, the matching procedure associates 2D camera coordinates to 3D world coordinates, yet the calibration is intended to produce the rotation matrix R_S^C and the translation vector T_S^C between CF and SF, therefore, after defining the pose of the Support Frame and

assuming it's fixed in relation to the WF, the calibration markers' coordinates P^W are transformed from world to support as follows:

$$P^S = P^W \cdot R_W^S + T_W^S \tag{4.4}$$

The function *estimateWorldCameraPose* employs a P3P+RANSAC algorithm to determine the pose parameters between CF and SF and eliminate outliers, if present, with a maximum reprojection error of 1 px and a confidence of 99%. The matched features detected throughout the calibration images are, then, fed to the optimization algorithm that adopts a Levenberg-Marquardt non-linear solution to the least-square problem. The procedure adopts the norm of the reprojection error to find local minima. The calibration results in an optimal calculation of the rotation matrix R_S^C and the translation vector T_S^C with a mean error of 0.55 px. The pose parameters are, then, concatenated with the transformation between SF and WF (**Eq. 4.5**) in order to obtain the true pose of the left camera frame in respect to the motion capture world reference.

$$R_W^C = R_S^C \cdot R_W^S$$

$$T_W^C = (T_S^C - T_S^W) \cdot R_W^S \tag{4.5}$$

Another calibration is required concerning the transformation matrix between the CubeSat Body Frame (BF) and the so-called Model Frame, identified by a combination of the six markers attached to the model structure. As previously described, three markers are positioned on the top face of the main spacecraft body while the remaining three are located on the base of the support frame (**Fig. 4.6**), whereas the Body Frame origin is located at the center of the bottom face of the mock-up and its axes are oriented as it's displayed in **Fig. 4.7** The main purpose of adopting multiple markers is to ensure that at least three points are always detectable and, therefore, it's always possible to define a triplet of linearly independent vectors as Model Frame, since occasionally a portion of them isn't visible by motion capture IR cameras or the ZED sensor. Besides, the top markers are positioned closer than the bottom ones and a longer baseline between markers improves the precision of the pose estimation.

The first step consists in elaborating and reconstructing the 3D coordinates of the markers from the motion capture data; the SPARTANS software associates tags with pre-defined detected features and tracks them to prevent the presence of outliers. Nonetheless, rarely the tracking technique fails and tagged elements are swapped with outliers, thus, the motion capture data need to be partially reconstructed and manually associated with the respective marker points. Concurrently, the markers' image coordinates must be identified and matched through the monocular camera. This task is accomplished by a dedicated MATLAB script which employs the same technique used with the calibration function. The marker detection is initialized by a combination of circular blob detection and nearest neighbour search to provide first guess correspondences, thereafter, a computer vision algorithm follows the 2D points using a Kanade-Lucas-Tomasi feature-tracking architecture. The tracker substitutes the circle detector whether a marker's view is occluded and it doesn't appear in camera or more markers overlap. The correspondences between image and world coordinates are fed to the P3P+RANSAC solver which outputs the relative pose between Model Frame and Camera Frame (R_M^C, T_M^C).

The successive step involves the determination of the rotation matrix R_M^B and translation vector T_M^B between BF and MF. The process is twofold: first, the pose estimation between the Body Frame and the Camera Frame (R_C^B, T_C^B) is performed for manually annotated features for a smaller set of images and secondly, together with the results of the previous calibration subsystem the following parameters are computed:

$$R_M^B = R_C^B \cdot R_M^C$$

$$T_M^B = (T_C^B - T_C^M) \cdot R_M^C \tag{4.6}$$

Assuming the rigidity of the mock-up and its support the pose parameters are constant throughout the set of images, consequently, a global optimization process is performed to reduce the 3D rms error of the projected mock-up points. The R_M^B computation and refinement functions as a non-linear optimization code that selects random subset of pose parameters and computes the mean value, thereafter, the set of transformation matrix with the lowest 3D rms error is chosen as optimal.

Given that the 3D coordinates of the wireframe model implemented in the SVD pose initialization architecture (**Fig. 4.7**) are expressed in the Body Frame but only the markers which are expressed in the Model Frame can be detected and tracked by the

motion capture, therefore it's necessary to explicate the roto-translation operation between the two reference systems.

Finally, the true pose of the satellite mock-up in respect to the camera frame is calculated by applying the following equations

$$R_C^B = R_M^B \cdot R_W^M \cdot R_W^C$$

$$T_C^B = (T_M^B - (T_W^C - T_W^M) \cdot R_M^W) \cdot R_C^M \tag{4.7}$$

where (R_M^W, T_M^W) are computed for different poses by the SPARTANS measurements system, (R_C^W, T_C^W) are constant since the camera is fixed and (R_M^B, T_M^B) has been calculated by the calibration process.

4.3 Dataset and model generation

The data collected with the SPARTANS facility are used to test the version of the SVD architecture for pose initialization under more restricted conditions than with the PRISMA SPEED dataset. Besides, the CubeSat mock-up presents similar elements to the Tango spacecraft, such as a large and easily identifiable solar panel structure and a simple polyhedral body with distinctive edges. The main difference between the two target vehicle is the number of antennas, their configuration and their length. The choice for the mock-up dynamics is to impose a simple rotation around the Z_{body} axis and tilting the base toward the source of light. The rotative stage turns at a velocity of 800 steps/s that translates to a rotation of 2° per second. The light source is positioned on the camera plane at the right of the stereo sensor so that the resulting scenario is a spinning satellite with variable light incidence. The relative distance between the mock-up and the camera is nearly constant and equal to about 1.65 m. The relative attitude, instead, varies slightly in relation to the camera reference system, in fact, considering that the axis of rotation is tilted at about 60° towards the light source. The proposed configuration allows the mock-up to show to the camera the larger polygonal tetrad, namely the solar panel, possibly enhancing the pose estimation success rate, as determined in the previous chapter.

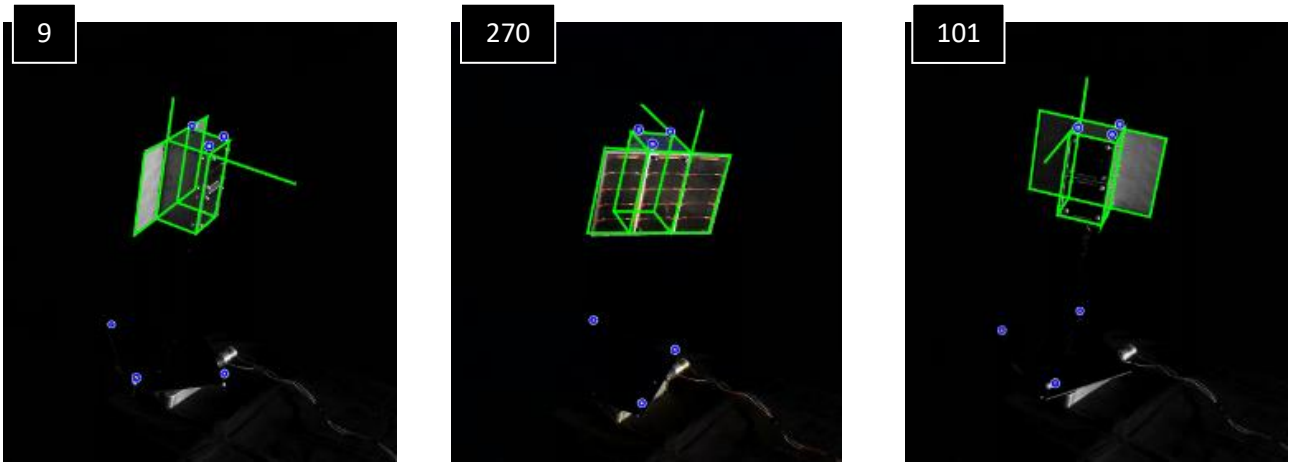


Figure 4.6 *The images are part of the dataset and show the CubeSat model in different stages of the rotation. In blue the markers detected by the calibration algorithm through the circular blob identification technique. The markers define a Model Frame which is correlated to the true pose of the target by means of a calibrated transformation matrix. In green it's shown the reprojection of the wireframe model of the true pose estimation and it can be noted how it fits the target's figure.*

The images are captured at a frame rate of 2 fps with high contrast and low ISO to reduce the background noise. A simple synchronization algorithm is implemented by comparing the start and finish time of the mock-up motion both for the markers' image coordinates and motion capture measurements so as to determine the true relative pose of the CubeSat mock-up.

The mock-up is built with simple elements: a 100x100x200 mm 3D- printed plastic structure is covered by 100x100 mm steel plates and aluminium angle brackets; moreover, two 100x100 mm plexiglass plates are positioned on the front face and a single 300x200 mm plexiglass plate is attached on the opposite surface to resemble an unfolded deployable solar panel; lastly, two 3 mm thick aluminium rods (100 mm and 200 mm) are employed to simulate the presence of antennas.

A wireframe model of the CubeSat is required to identify the 3D high-level features needed to perform the pose initialization. The geometric model is built following the same indications contained in the paper by Sharma et al. **[31]**, i.e., reducing the complexity of the geometry without sacrificing the fidelity of the model. The CubeSat mock-up is already a simplified version of a real satellite, yet, the wireframe model presented in **Fig. 4.7** can be considered a sufficiently accurate schematic representation of a 2U-CubeSat.

Among the linear elements of the CubeSat wireframe model, the following high-level 3D features have been defined:

- 7 polygonal tetrads
- 9 polygonal triads
- 23 parallel triads
- 24 parallel pairs
- 20 proximal pairs
- 2 antennas

It's important to notice how the CubeSat mock-up presents a larger number of comple geometrical groups and less simple ones in respect to the Tango model, and there is a larger set of symmetrical combinations of edges that convey indetermination to the pose solution problem.

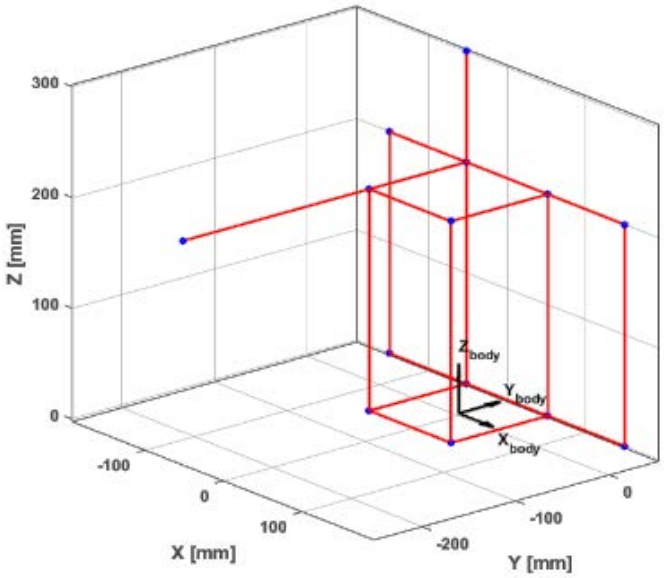


Figure 4.7 *Wireframe model of the CubeSat mock-up*

4.4 Algorithm implementation

The SVD algorithm is a model-based pose initialization, therefore, it requires model-specific settings and implementation for each target vehicle and, in the case of the proposed experimental setup, the algorithm needed to be tailored to serve the scope of correctly determining the relative pose. The main differences between the SVD implementation previously described and the one used for the SPEED data are presented below.

Firstly, the images captured by the ZED left camera require a preliminary cropping process to eliminate the background interference of laboratory equipment outside the scene. A rectangular mask of 710×450 px is defined and employed to restrict the area of application of the image processing subsystem.

The Weak Gradient Elimination technique for ROI detection is implemented and, given the absence of Earth-like simulated background for the dataset, a wider range for gradient elimination is chosen. The exponential fitting of the gradient distribution function computed with the Prewitt operator, is limited by a threshold of 99% and the weakest edges are eliminated. The respective cumulative distribution is limited to the central 99% of the gradient's incremental sum along the axes and, as a consequence, the limits of the target's bounding box.

The edge detection is accomplished by merging the two streams of data obtained by the ROI+Hough process and the Sobel+Hough technique, as with SPEED. The Hough transform is applied to the image gradient after an ulterior prefiltering that eliminates the edges with an intensity inferior to 0.15 (normalized values). The resolution set for the edge detection remains equal to $\rho = 0.5\ px$ and $\theta = 0.1°$ but the threshold for peaks is lowered to 0.001. As in **Eq. 3.2**, the geometric limits for line identification are set as follows:

$$l_{min}^{H} = 0.15 \cdot l_{ROI} \qquad \lambda^{H} = 0.02 \cdot l_{ROI} \tag{4.8}$$

The minimum length of the detected segments is greater than for Tango, considering the proportion among the shorter edges and the characteristic length of the wireframe model. A merging procedure within lines is implemented to filter shorter fragmented elements through the application of an angular threshold of 10° and a radial threshold of $0.1 \cdot l_{ROI}$.

As regards the Sobel+Hough stream of data, the process for identifying the true edges relies on the same architecture described in the previous chapter. A lower threshold for the Sobel operator is set, equal to 0.03, while the Hough transform hyperparameters are set as follows:

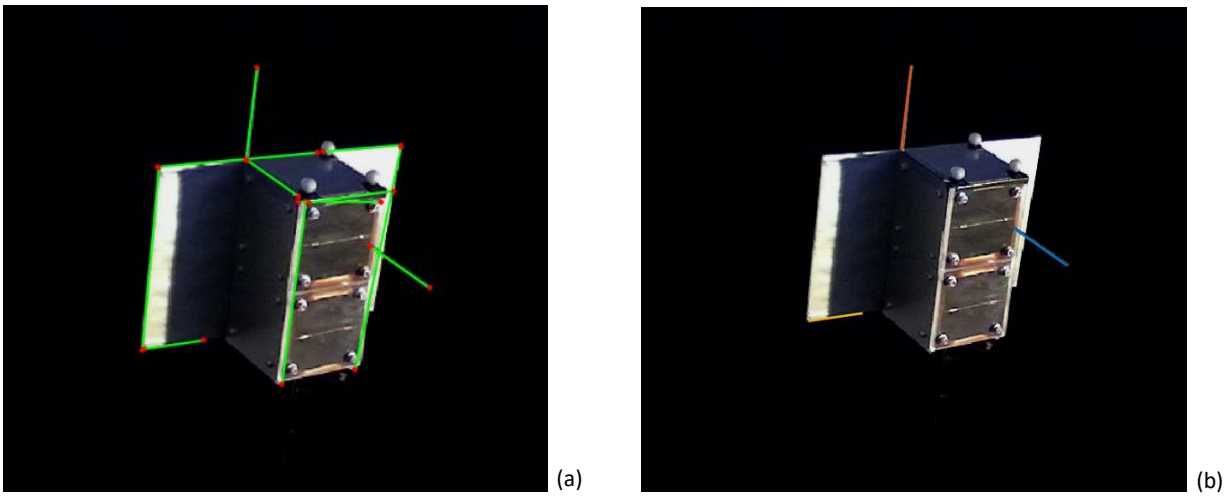
$$l_{min}^{H} = 0.1 \cdot l_{ROI} \qquad \lambda^{H} = 0.02 \cdot l_{ROI} \tag{4.9}$$

The detected edges from the two streams of data are merged, or eliminated if overlapped, in relation to midpoint distances, orientation and radius distance. The merging process is performed by a dedicated MATLAB script.

Consecutively, the segments are fed to the perceptual grouping block which arranges them in high-level features. The 2D geometric conditions are similar to the one employed for SPEED except for the antenna identification. The mock-up antennas can't be discriminated on the basis of length in respect to the remaining edges, consequently, alternative methods must be implemented. In fact, the limited variation of attitude represented in the dataset and the absence of disturbing elements in the background allow to impose the proximity to ROI limits as discriminating condition for antenna identification. Practically, distance between detected edges' endpoints and ROI limits is tested and a threshold equal to $0.05 \cdot l_{ROI}$ is imposed to discriminate the features. Despite this condition, spurious edges are still classified as antennas, hence an evaluation system is devised to rate the segments that meet the first condition. Assuming that the antennas' extremal points are farther away from the structure than the remaining points, the candidate features are rated on the basis of the sum of the relative Euclidean distance between themselves. This method associates the lower sum value to the points and edges that belong to the CubeSat main body. Only the two most high-rated edges are classified as antennas.

In addition to the feature synthesis subsystem, the matching procedure it's also modified due to the lack of asymmetrical elements in the wireframe target model. The repetition of similar edges produces multiple identical 3D parallel pairs and triads, as well as proximal pairs and polygonal tetrads. Therefore, the presence of distinguishable elements like oriented edges (antennas) plays a fundamental role in the effectiveness of the pose initialization. A mitigation to the problem comes from adding the possibility of considering the combination of parallel triads with antennas. Parallel triads are easily and commonly detected in the case of the CubeSat mock-up and, although they are sufficient to provide a pose estimation, there are multiple correct solutions to the Perspective-n-point problem; only pairing the triad with an antenna can reduce the solution space. An example of the results of this technique is shown in **Fig. 4.8** below.

Finally, the pose determination and refinement tasks are accomplished by a P3P algorithm followed by a non-linear Levenberg-Marquardt optimization, as with the SPEED data.



(a) (b)

Figure 4.8 *In (a) the lines obtained through the merging of the detected edges, in (b) the segments identified as antennas. One of the antennas is correctly classified while the other two are respectively a partial edge and an incorrectly identified segment.*

4.5 Validation results

The dataset comprises of 428 images of the CubeSat mock-up and it's fed to the SVD pose initialization algorithm. The overall performance parameters are summarized in **Tab. 4.3**.

Success Rate	37.12 %
$\%_{Pose}$	96.96 %
Mean IoU	77.39 %
Mean E_T [mm]	23.02
Mean E_R [deg]	3.59
Mean ERR_T [mm]	[5.90 4.94 20.34]
Mean ERR_R [deg]	[1.02 2.28 1.96]
Median N_{lines}	4
Median N_{points}	11
Mean N_{comb}	5556
Mean t_{comp} [s]	11.2604

Table 4.3 SVD architecture performance metrics. The accuracy values refer to high-confidence solutions only, while the remaining parameters are calculated over the entire dataset.

Clearly the SVD method proves to be consistently more effective when applied in a more limited range of conditions. The success rate it's five time greater than the one observed with the SPEED data and 13 out of 428 images resulted in a ROI-only estimation due to an insufficient number of edges or high-level features. The algorithm produces a higher number of pose solution (N_{pose}) over the totality of data with improved ROI detection outputs varying between a minimum of 50 % to a maximum of 90 % of overlapping area (IoU). A correct identification of the target's figure bounding box, as previously pointed out, affects positively the number of true line correspondences that, in fact, amounts to a mean value of four correct edges. Along with a higher N_{lines}, the number of correspondences between true 3D points and detected endpoints is equal to 11, notifying the increase in the edge detection process's effectiveness. Considering percentage values, on average 45 % of the detected segments match the model's true edges and 68 % of their endpoints with the true vertices of the CubeSat within the limit of a threshold distance equal to $0.025 \cdot l_{ROI}$.

The rotational and translational accuracy of high-confidence solutions is comparable with the error achieved with SPEED's validation and in accordance with the results presented in the SVD paper **[31]**. In particular, the translational error is greater along the camera's boresight axis Z_{CAM}, as the range in this direction is not directly related to the detected lines, in contrast to the position determination along the other directions of the camera frame. Nonetheless, the mean error E_T is lower than the one registered with SPEED and this can be easily explained given the fact that the relative distance between camera and target is constant and inferior to 3 m. As shown in **Fig. 3.12**, a shorter relative range improves the accuracy of the estimation reducing steadily the translational error, hence, accordingly, E_T is equal to 2.3 cm at about 1.65 m. The rotational precision is, instead, slightly worse than that resulted from the previous validation test, even though still lower than 5°. A little bias in the measurement's error can be attributable to the simulated illumination condition of the SPARTANS experimental setup. The orientation precision is related, in part, to the sharpness of the contrast gradient along the satellite edges, assuming that, as it is for the two validation dataset, the Hough transform operates with the same radial and angular resolution in the edge detection process. Some elements affect the fidelity of the simulated laboratory illumination: (1) the light source consists in a lamplight at short distance so the light rays can't be considered parallel as would happen with solar light in a LEO orbit, (2) the black cloth is obviously more reflective than the space background and (3) the real edges of the mock-up have an intrinsic uncertainty related to the build quality.

Furthermore, the number of combination and the computational time are greater than in previous results but comparable due to the fact that the number of elements belonging to the 3D high-level features is greater for the CubeSat model than Tango's model, thus the number of possible combinations to test for pose determination directly increases.

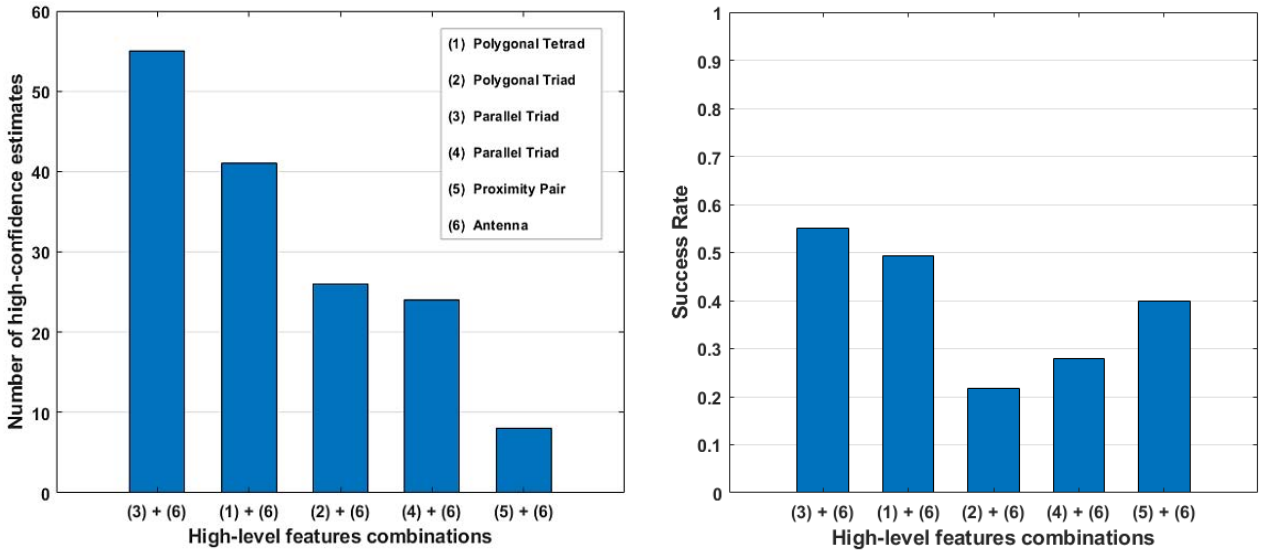


Figure 4.9 *Distribution of high-confidence pose solution in relation to the feature group combinations and the respective success rate.*

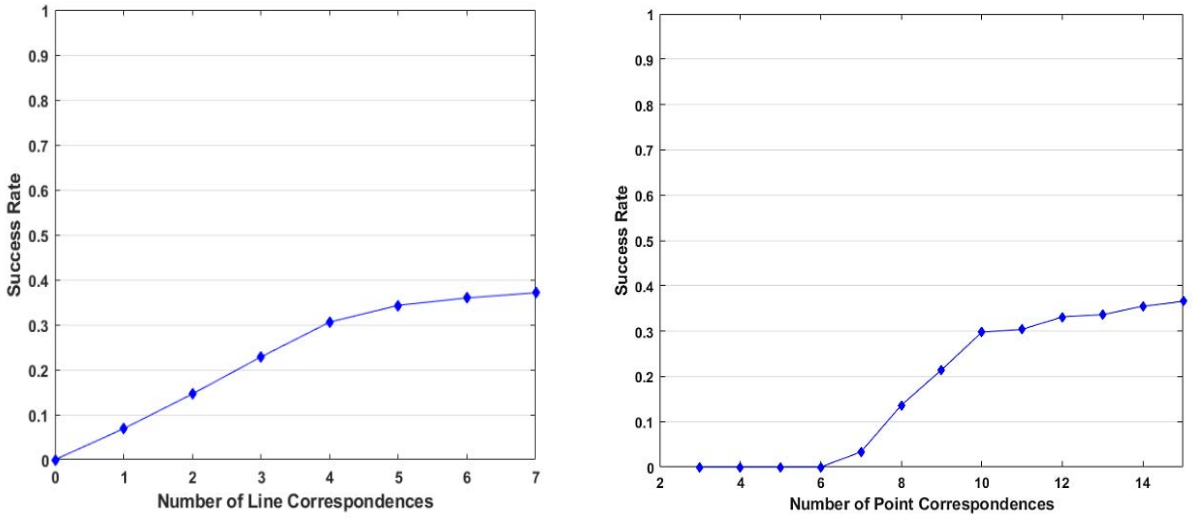


Figure 4.10 *Success rate in relation to the number of line and point true correspondences.*

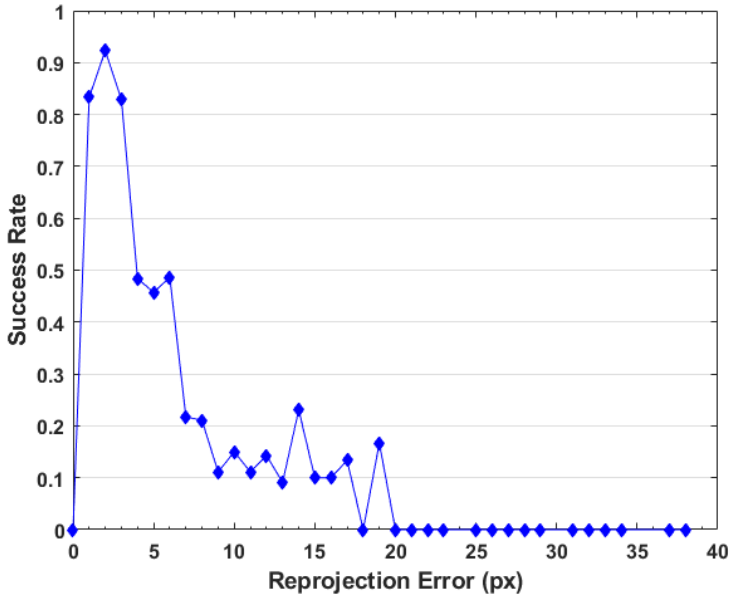


Figure 4.11 *Plot of the success rate against the 2D reprojection error of the pose outputs.*

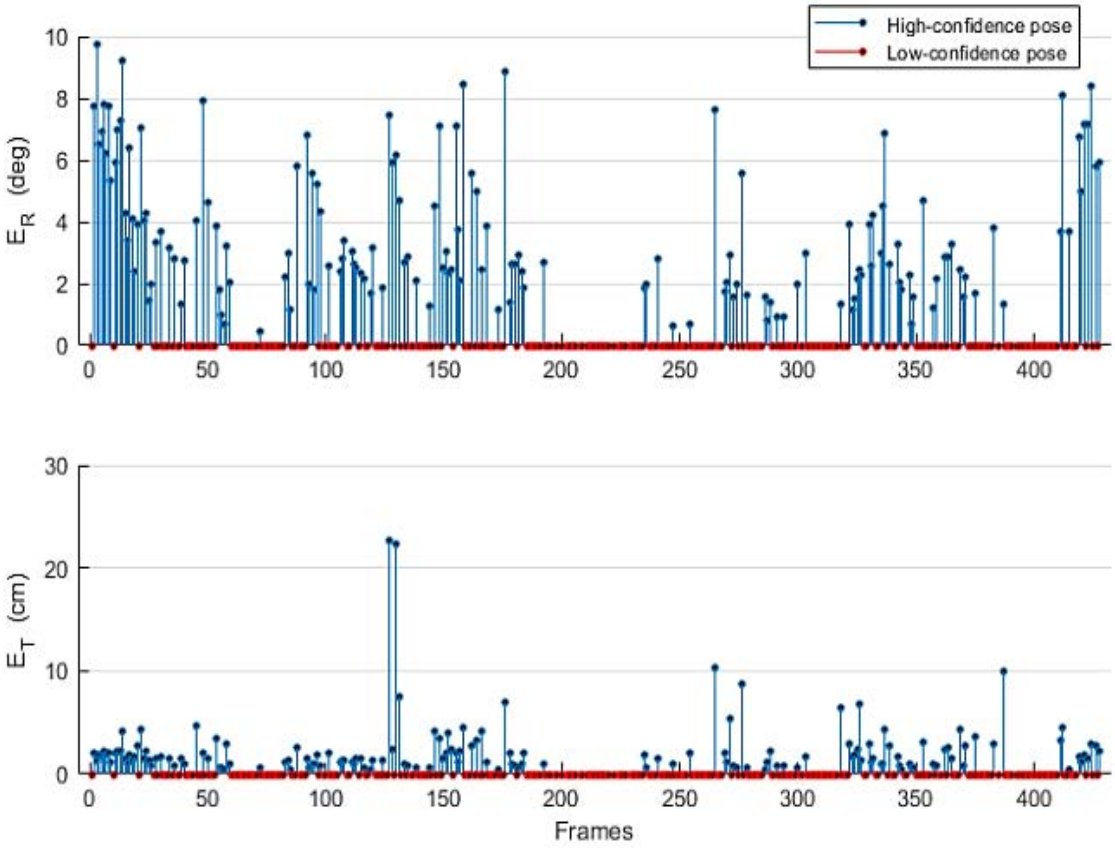


Figure 4.12 *Distribution of high-confidence and low-confidence pose solutions across the recorded frames of the CubeSat dataset. The correct estimates are plotted with their relative translational and rotational accuracy.*

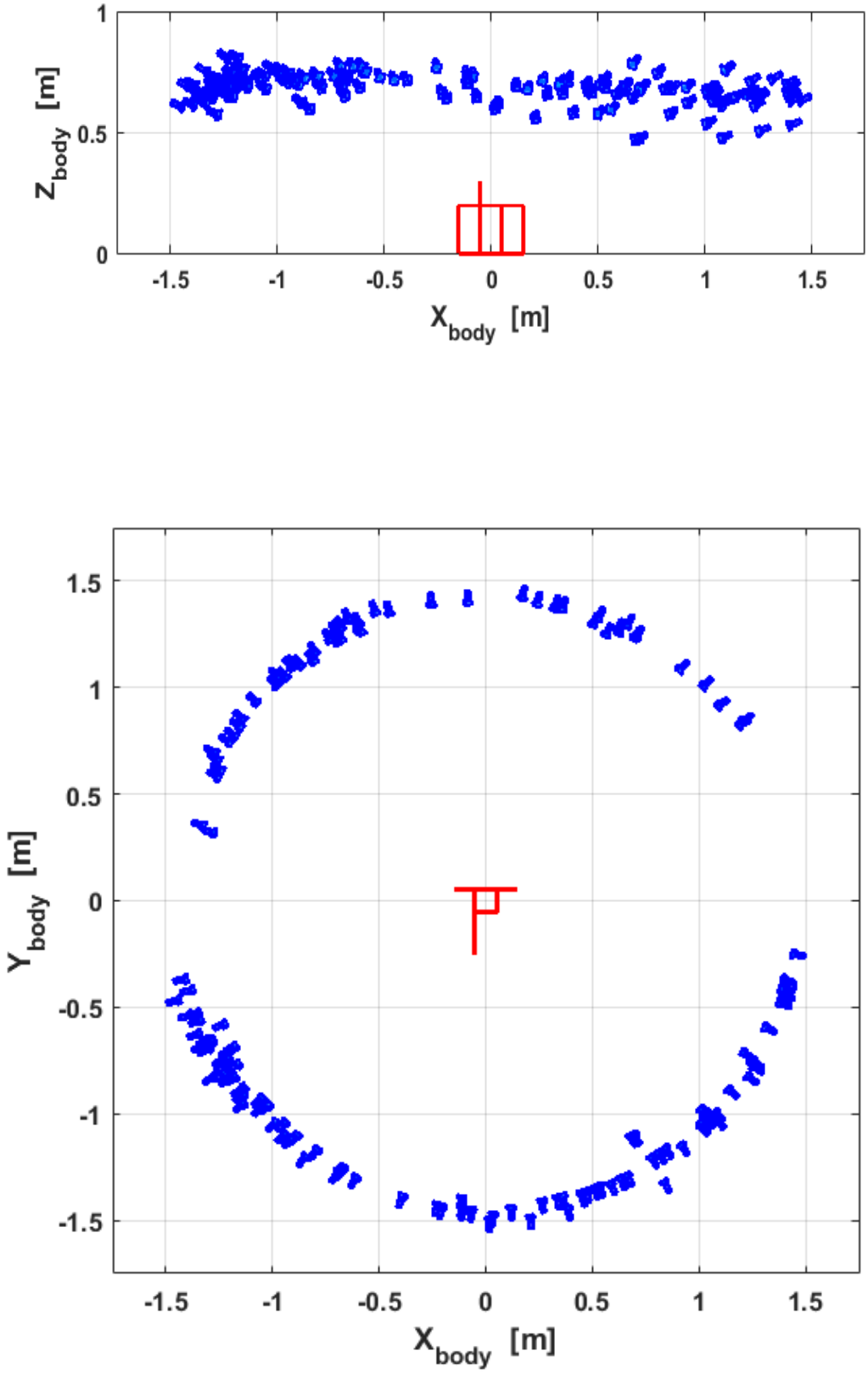


Figure 4.13 *Spatial distribution of high-confidence pose estimates.*

The **Fig. 4.9** shows the distribution of high-confidence pose solutions and the pairs of complex and simple geometric features fed to the pose solver. The plot demonstrates that combinations which produce correct poses always involve at least an antenna feature, since it discriminates uniquely among symmetrically valid poses in the solution space. The most common combination is the coupling of parallel triads and antennas and the reason why it exceeds in frequency and effectiveness the pair composed by polygonal tetrads and antennas, which is successful for the Tango vehicle, it's the fact that the CubeSat model presents numerous triads of parallel edges. Let it be noted that this particular combination results in the highest number of possible matches to subject to the pose solver and, consequently, it requires a longer computational time. Nonetheless, these combinations provide a success rate equal to 54,34 %, which is consistently above the average value registered in other validation tests.

In **Fig. 4.10** the success rate is correlated to the number of line and point true correspondences and, as with the SPEED data, it's clear how the pose solving algorithm outputs more correct poses with the increase of true edge detection. Despite this, given the nature of the perceptual grouping technique and the P3P algorithm for pose solving the total efficiency of the SVD method reaches an intrinsic limit at about 40 %.

In addition to the previous plots, the figure in **Fig. 4.11** displays the variance of success rate in relation to the reprojection error E_{2D}. Since the precision of the pose estimation is correctly evaluated by computing the absolute difference between estimated and true translation vectors and rotation matrices, the reprojection error isn't a trustworthy performance metric but can act as a preliminary metric for the continuation of the algorithm. As shown in the plot, values of reprojection error inferior to 5 px entail a higher success rate but as E_{2D} increases the success rate drops rapidly. On the basis of this observation, in order to improve the overall effectiveness of the SVD pose initialization within a navigation architecture, a threshold for the reprojection error associated with the best pose solutions can be set to break the process and reinitialize the pose estimation.

In **Fig. 4.12** and **Fig. 4.13** the high-confidence solutions spatial distribution and their respective rotational and translational errors are depicted. It can be easily noted how there are clear gaps in the distribution and those gaps are correlated to the sides of the CubeSat mock-up with the lowest visible area. As it has been noted in the previous chapter with the Tango spacecraft, the visibility of clear and easily identifiable elements, like large solar panel, ensures a higher probability of resolving exact relative poses. As shown in **Fig. 4.14**, the CubeSat model is depicted from a point of view that captures the main solar panel. Given the presence of the panel pattern, a

part of the lines that are detected have no correspondence with the wireframe model, yet the perceptual organization algorithm provides the correct identification of a polygonal triad and an antenna, thus computing the right pose. This example demonstrates how the algorithm can overcome the presence of spurious or useless edges by applying the feature synthesis technique.

Lastly, **Fig. 4.15** collects some of the results of the SVD method for pose initialization, both high-confidence, low-confidence and ROI-only solutions. In the case of the images 2, 335, 145 the pose initialization is effectively producing correct outputs, while the results in images 211 and 406 are wrong. In particular, the pose solution for the image 406 is incorrectly estimated due to multiplicity of solutions based on the same match of an antenna and a parallel triad. The image 22 shows how the Weak Gradient Elimination technique operates optimally with the illumination and background conditions imposed.

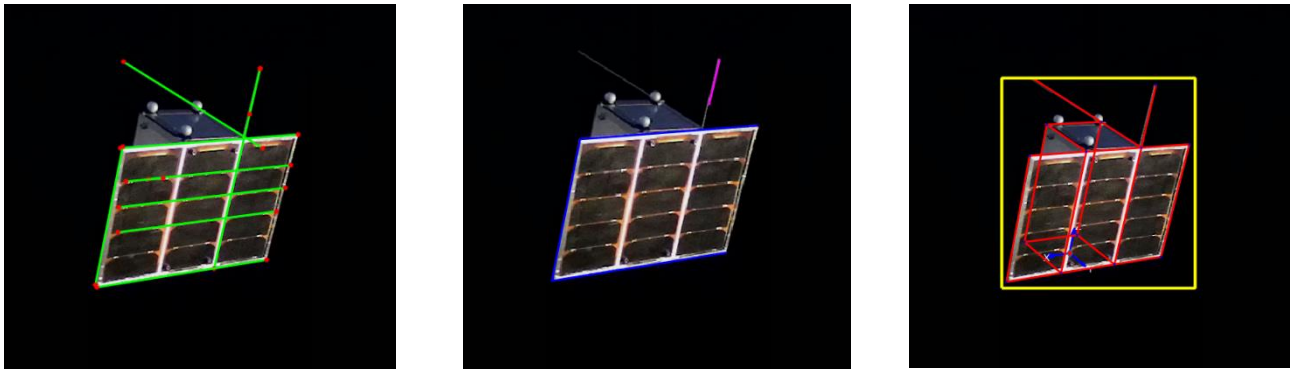


Figure 4.14 *Results of algorithm subsystems for the frame 46 of the dataset. From left to right, the detected lines, the couple of polygonal triad and antennas that produce the optimal solution and the model projected on the image after correct pose estimation.*

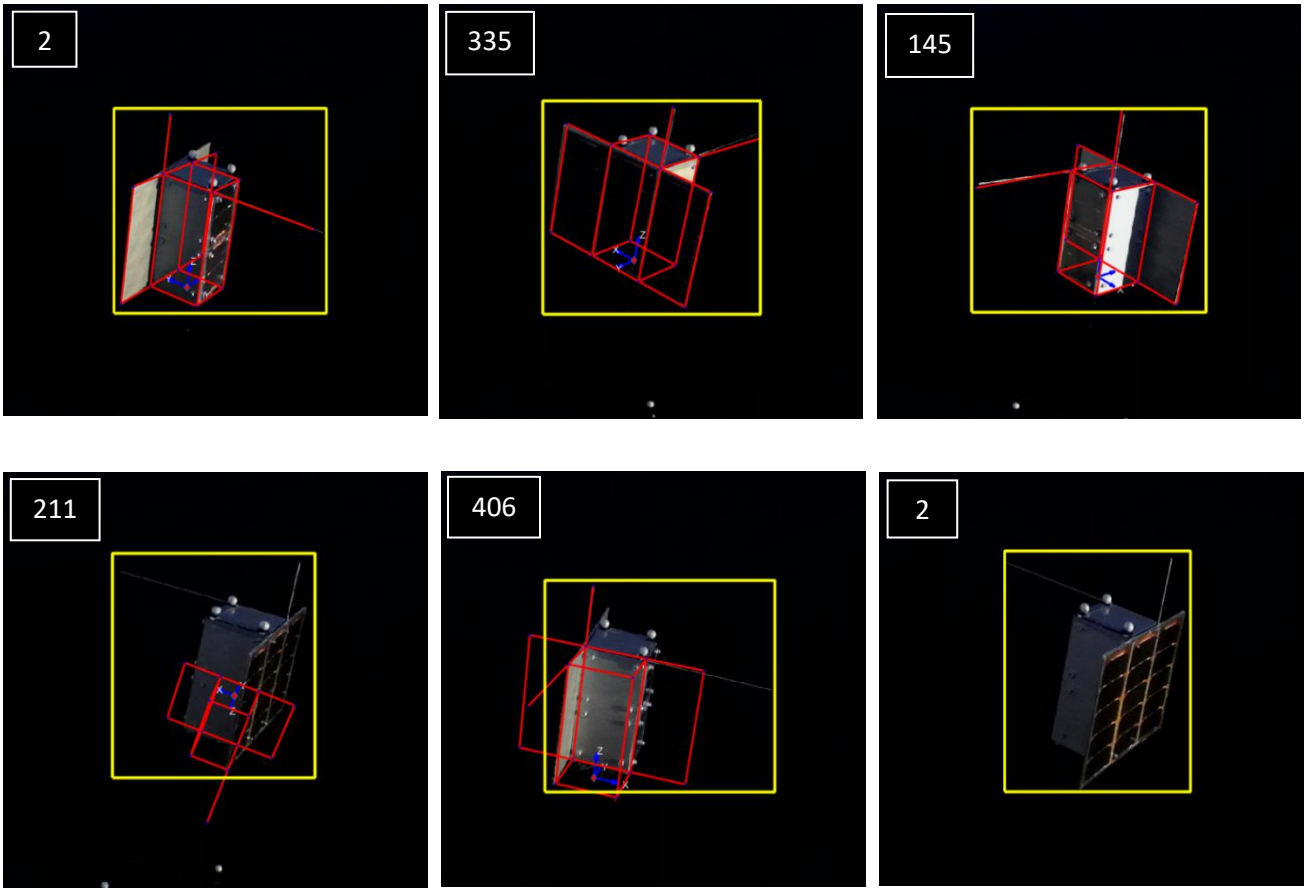


Figure 4.15 *Some results from the application of the SVD method to the CubeSat dataset.*

Conclusions

The SVD architecture for pose initialization, proposed in **[31]**, whose validation tests are presented in this thesis work, has proven to be sufficiently accurate only in a limited operational range in terms of relative distance, model geometry and environmental conditions. The SVD method addresses the problem of determining the relative pose of non-cooperative passive space vehicle, i.e., target spacecraft that are not able to establish a communication link with the chaser and it's not equipped with known visual markers. The scenario of non-cooperative passive targets occurs frequently in active debris removal (ADR) and on-orbit servicing (OOS) missions which involve uncontrolled or defunct spacecrafts. The technique consists in a model-based feature matching pose initialization process which employs single low-resolution images captured by a monocular vision system and a known, simplified, geometric 3D model of the target to produce, firstly, a coarse measurement of relative range, line of sight and, secondly, an accurate estimate of relative orientation and position without a priori information. The technique brings four fundamental innovations in the field of monocular feature-based pose estimator:

- A Weak Gradient Elimination (WGE) procedure is introduced to define the target's region of interest in the image.
- The hyperparameters of the Hough transform are expressed as scalars of the ROI's dimensions so to provide a more flexible and efficient edge detection.
- The detected segments are organized and classified in high-level geometric features on the basis of endpoints proximity and parallelism
- The search space for pose determination is greatly reduced by combining the most complex and the simplest high-level features detected

After providing an initial overview on the issues and the state-of-the-art instruments employed to determine the relative pose in various operative scenarios and highlighting the advantages of the monocular approach, a detailed description of the SVD technique is proposed, together with the conclusions drawn in the paper **[31].** The paper demonstrates the effectiveness of the SVD method with two datasets of 25 and 142 of PRISMA mission's Tango spacecraft real images. The results of multiple

validation tests prove separately that (1) the gradient-based ROI identification outperforms typical blob detectors in terms of accuracy and computational time, (2) the combination of Hough transform and WGE produces a higher positive rate of true feature correspondences than other edge and feature extractors without requiring a great computational cost and (3) the SVD method outputs fewer pose estimates than a simple RANSAC procedure but the high-confidence results are consistently more accurate than the one provided by random sample consensus. In regard to the PRISMA image set the paper shows that only half of the attempt result in a pose solution and only 5 out of 12 estimates are correct with mean accuracies of 1.59° and 0.53 m.

The main contribution of the thesis is to implement the SVD architecture within the MATLAB software and validate it with a larger dataset of synthetic imagery of the Tango vehicle and, subsequently, developing a version of the algorithm based on a different vehicle and testing the efficiency in that case. Some modifications are considered in the specific implementation of the SVD method used in the validation tests. Besides the particular characterization of the Hough hyperparameters and the addition of geometric conditions for feature organization, the proposed pose solving technique, i.e., a EPnP solver followed by a simple Newton-Raphson optimization, is discarded and a P3P + RANSAC solver followed by a non-linear Levenberg-Marquardt optimization. This choice is based on the observation that the number of points to solve the perspective equation is always limited to a maximum of six, given the coupling of two geometric feature, and that EPnP is greatly affected by outliers and collinear points, commonly detected throughout the dataset.

The synthetic imagery of the SPEED database consists in 12000 frames of a computer-generated Tango model with a statistical distribution of attitude and position values, as well as a series of cropped Earth background. The application of the SVD algorithm has shown its intrinsic limitations with the SPEED data. In fact, the percentage of high-confidence solutions it's reduced to 8% and about 40% of the attempts results in ROI-only measurements. The validation shows the dependency of the feature matching and classification effectiveness on the correct identification of the target's ROI, which in turn tends to be incorrect in the presence of composite background. Nonetheless, accurate pose solutions are estimated with a mean precision of 2.85° and 0.09 m, which is comparable to the results in **[31]**. In particular, the accuracy decreases in accordance with the increase of relative distance and it's predominantly distributed along the boresight axis of the camera. The SPEED validation proves that the operative range of the SVD method is between 3 m (the minimum relative distance of the dataset) and about 15 m. The analysis of high-confidence solutions' spatial distribution and the frequency of feature combinations

provides a fundamental outcome, i.e., the technique is considerably more effective when a large polygonal tetrad, usually a easily identifiable solar panel, and an antenna are visible. This proves the basic assumption that the search space for pose determination can reasonably be reduced by combining a complex and a simple high-level feature, yet the method provides a sufficiently robust performance only when those features are uniquely identifiable and asymmetrically oriented in relation to the target's body frame.

The results of the first validation test are expanded and confirmed by the second test which involves a CubeSat mock-up mounted on a rotative stage. The dataset is acquired at the SPARTANS testbed for relative satellite dynamics and consists in a set of 428 images and respective true reference pose values. The CubeSat dataset presents more limited conditions such as constant illumination, empty black background and close-range operativity (relative distance of ~1.65 m). The results show that the absence of background interference translates directly to a higher accuracy of the ROI detection process, which, as a consequence, increases the mean number of true line correspondences and the success rate of the poses solver. The mean success rate is equal to 37 % and the mean error for the high confidence solutions have an accuracy of 0.02 m and 3.59°. Let it be noted that each target model entails a particular number of 3D features, for instance, the most frequent combination in the case of the CubeSat mock-up is the parallel triad, yet for this reason, it implies the possibility of symmetrical solution if not paired with a discriminating element like an antenna.

The validation tests address the main limitations of the SVD technique. While resulting in sufficiently accurate estimations for the initialization process of a relative guidance and navigation system, it lacks an adequate robustness to the variety of scenarios present in complex future missions that involve non-cooperative passive spacecrafts. The algorithm effectiveness is highly dependent on the geometric properties and on the intrinsic complexity of the target's structure. The visual recognizability of certain geometric features like polygonal structure and antennas, in combination with the asymmetry of their distribution and the detectability of their edges, improves consistently the success rate of the algorithm up to 50% of high-confidence solutions. Regardless of the efficiency, the estimation procedure actually reduces the solution search space, thus, allowing the implementation of on-board executable and fast software, required by small operative satellites during autonomous navigation. Above all, it's fundamental to highlight how the SVD architecture can be optimized by setting breakpoints correlated to the number of detected edges or the reprojection error in order to re-initialize the algorithm when certain conditions that yield a lower success rate are met. Besides, the pose estimates

provided by SVD represent first attempts to compute the relative pose of the target and, if integrated with a far- range angles-only navigation system and a pose tracking algorithm, it can effectively reduce the uncertainty of the estimation.

In conclusion, it's worth noticing that the SVD architecture represents an attempt to introduce an intelligent design for the feature detection and matching process within the classical Perspective-*n*-Points problem thereby producing a consistent improvement in terms of accuracy and computational time. Following this path, the use of convolutional neural networks and machine learning aims at overcoming the fundamental limitations of point-based methods and providing surprising accuracy and speed combined with an extreme operational flexibility.

Acknowledgments

I would like to thank my thesis advisors Prof. Stefano Debei and Prof. Marco Pertile that allowed me to pursue my interest in the topic of the thesis and provided me with useful advices and fundamental suggestions. In addition, I would like to thank Sebastiano Chiodini, Andrea Valmorbida and Mattia Mazzucato that were involved in the creation and preparation of the experimental setup and helped me with the acquisitions at the SPARTANS facility.

Finally, my gratitude goes to my family for supporting me throughout these years of study and always providing encouragement and inspiration in my moments of doubt.

References

[1] Zimpfer, D., Kachmar, P., and Tuohy, S., *"Autonomous Rendezvous, Capture and In-Space Assembly: Past, Present and Future"* 1st AIAA Space Exploration Conference: Continuing the Voyage of Discovery, NASA Johnson Space Center, Orlando, FL, 2005. doi: 10.2514/6.2005-2523

[2] Starek, J.A., Acikmese, B., Nesnas, I. A. D. and Pavone , M., *"Spacecraft Autonomy Challenges for Next Generation Space Missions"*, Lecture Notes in Control and Information Sciences, Ed. E. Feron, Vol. 460. Springer Berlin Heidelberg, 2016, pp. 148.

[3] Ventura, J., *Autonomous Proximity Operations for Noncooperative Space Target* , PhD, Technische Universität München, Munich, DE, 2016.

[4] Opromolla, R., Fasano, G., Rufino, G., Grassi M., *"A review of cooperative and uncooperative spacecraft pose determination techniques for close-proximity operations"*, University of Naples "Federico II", Dept. of Industrial Engineering, P.le Tecchio 80, 80125, Napoli, Italy.

[5] R.T. Howard, T. C. Bryan, *"DART AVGS Performance"* Proceedings of SPIE, 'Sensors and Systems for Space Applications,' Richard T. Howard, Robert D. Richards, Editors, 65550L, Vol. 6555, May. 3, 2007.

[6] T. M. Davis, T. L. Baker, T. L. Belchak, W. R. Larsen, *"XSS-10 Microsatellite Flight Demonstration Program,"* Proceedings of AIAA/USU Conference on Small Satellites, Logan, UT, USA, Aug. 11-14, 2003.

[7] R. Richards, J. Tripp, S. Pashin, D. King, J. Bolger, M. Nimelman, *"Advances in Automous Orbital Rendezvous Technology: The XSS-11 Lidar Sensor,"* Proceedings of the 57th IAC/IAF/IAA (International Astronautical Congress), Valencia, Spain, Oct. 2-6, 2006.

[8] S. Persson, B. Jacobsson, *"PRISMA - Swedish in-orbit testbed for rendezvous and formation flying,"* Proceedings of the 57th IAC/IAF/IAA (International Astronautical Congress), Valencia, Spain, Oct. 2-6, 2006, IAC-06-D1.2.02

[9] S. D'Amico, E. Gill, O. Montenbruck, *"Relative Orbit Control Design for the PRISMA Formation Flying Mission,"* AIAA Guidance, Navigation, and Control Conference and Exhibit, Keystone, CO, USA, Aug. 21-24, 2006

[10] J. Harr, M. Delpech, L. Lestarquit, D. Seguela, *"RF Metrology Validation and Formation Flying Demonstration by Small Satellites,"* The 4S Symposium: Small Satellite Systems and Services, Chia Laguna Sardinia, Italy, Sept. 25-29, 2006.

[11] J. A. Christian, S. Cryan, *"A Survey of LIDAR Technology and its Use in Spacecraft Relative Navigation"* , AIAA Guidance, Navigation and Control Conference; Boston, MA; United States; 19-22 Aug. 2013.

[12] D.C. Woffinden, D.K. Geller, *Navigating the road to autonomous orbital rendezvous,* J. Spacecr. Rockets 44 (4) (2007) 898–909.

[13] A. Renga, M. Grassi, U. Tancredi, *Relative navigation in LEO by carrier-phase differential GPS with intersatellite ranging augmentation*, Int. J. Aerosp. Eng. 2013 (2013), 627509, 11 pages.

[14] S. D'Amico, J.S. Ardaens, S. De Florio, *Autonomous formation flying based on GPS—PRISMA flight results*, Acta Astronaut. 82 (1) (2013) 69–79.

[15] M. Delpech, J. C. Berges, T. Karlsson, F. Malbet, *"Results of PRISMA/FFIORD Extended Mission and Applicability to future Formation Flying and Active Debris Removal Missions,"* Proceedings of the 5th International Conference on Spacecraft Formation Flying Missions and Technologies (SFFMT), Munich, Germany, May 29-31, 2013.

[16] John L Junkins, Declan C Hughes, Karim P Wazni, and Vatee Pariyapong. *Vision-based navigation for rendezvous, docking and proximity operations*. In 22nd Annual AAS Guidance and Control Conference, Breckenridge, CO, pages 99–021, 1999

[17] NK Philip and MR Ananthasayanam. *Relative position and attitude estimation and control schemes for the final phase of an autonomous docking mission of spacecraft*. Acta Astronautica, 52(7):511–522, 2003.

[18] Chi-Chang J Ho and N Harris McClamroch. *Automatic spacecraft docking using computer vision-based guidance and control techniques.* Journal of guidance, control, and dynamics, 16(2):281–288, 1993.

[19] L. Mullins, A.F. Heaton, J. Lomas, *Advanced Video Guidance Sensor Inverse Perspective Algorithm*, MSFC Internal Memo TD45 (04–09), Dec. 2003.

[20] J. LeCroy, D. Hallmark, P. Scott, R.T. Howard, *Comparison of navigation solutions for autonomous spacecraft from multiple sensor systems*, in: Proceedings of the 2008 SPIE Defense and Security Symposium, Apr. 2008.

[21] R.T. Howard, A.F. Heaton, R.M. Pinson, C.L. Carrington, J.E. Lee, T.C. Bryan, B.A. Robertson, S.H. Spencer, J.E. Johnson, T*he advanced Video guidance sensor: orbital express and the next generation,* in: AIP Conference Proceedings, vol. 969, January 2008, pp. 717–724.

[22] Lance B Gatrell, William A Hoff, and Cheryl W Sklair. *Robust image features: Concentric contrasting circles and their image extraction*. In Robotics-DL tentative, pages 235–244. International Society for Optics and Photonics, 1992.

[23] Brent Tweddle. *Computer vision based navigation for spacecraft proximity operations.* S.M. Thesis MIT: SSL# 1-10, February 2010.

[24] Tweddle, B.E.; Saenz-Otero*, A Relative Computer Vision-Based Navigation for Small Inspection Spacecraft*. J. Guid. Control Dyn. 2014, 38, 969–978.

[25] G. Fasano, D. Accardo, M. Grassi, *A Stereo-vision Based System for Autonomous Navigation of an In-orbit Servicing Platform*, 2009. ISBN-10: 1-56347-971-0, AIAA Infotech@Aero-space 2009, Seattle, USA.

[26] S.B. Robinson, J.A. Christian, *Pattern design for 3D point matching*, Navigation 62 (3) (2015) 189–203

[27] D. Conte, P. Foggia, C. Sansone, M. Vento, *Thirty years of graph matching in pattern recognition*, Int. J. Pattern Recognit. Artif. Intell. 18 (3) (2004) 265–298

[28] P. David, D.F. DeMenthon, R. Duraiswami, H. Samet, *SoftPOSIT: simultaneous pose and correspondence determination*, Int. J. Comput. Vis. 59 (3) (2004) 259–284

[29] M.A. Fischler, R.C. Bolles, *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*, Commun. ACM 24 (6) (1981) 381–395.

[30] N.N. Dawoud, B.B. Samir, J. Janier, *Fast template matching method based optimized sum of absolute difference algorithm for face localization*, Int. J. Comput. Appl. 18 (2011) 30–34.

[31] S. Sharma, J. Ventura, S. D'Amico, "Robust Model-Based Monocular Pose Initialization for Noncooperative Spacecraft Rendezvous", Journal of Spacecraft and Rockets. 55. 1-16. 10.2514/1.A34124.

[32] S. Augenstein, S.M. Rock, *Improved frame-to-frame pose tracking during visiononly SLAM/SFM with a tumbling target*, in: Proceedings of the IEEE International Conference on Robotics and Automation, May 2011.

[33] B.E. Tweddle, *Computer Vision-based Localization and Mapping of an Unknown, Uncooperative and Spinning Target for Spacecraft Proximity Operations*, PhD, Massachusetts Institute of Technology, Cambridge, MA, 2013.

[34] Rosten, E., and Drummond, T., *"Machine Learning for High-Speed Corner Detection," Computer Vision–ECCV 2006*, Vol. 3951, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer, New York, 2006, pp. 430–443.

[35] Harris, C., and M. Stephens, *"A Combined Corner and Edge Detector,"* Proceedings of the 4th Alvey Vision Conference, August 1988, pp. 147-151.

[36] Leutenegger, S., Chli, M., and Siegwart, R. Y., *"BRISK: Binary Robust Invariant Scalable Keypoints,"* Proceedings of the IEEE International Conference on Computer Vision, IEEE Publ., Piscataway, NJ, pp. 2548–2555.

[37] Canny, J., *"A Computational Approach to Edge Detection*," IEEE Transactions on Pattern Analysis and Machine Intelligence,Vol.PAMI-8, No. 6, 1986, pp. 679–698.

[38] Duda, R. O., and Hart, P. E*., "Use of the Hough Transform to Detect Lines and Cures in Pictures,"* Communications of the Association Computing Machinery, Vol. 15, No. 1, 1972, pp. 11–15.

[39] Prewitt, J*., "Object Enhancement and Extraction,"* Picture Processing and Psychopictorics, Vol. 10, No. 1, 1970, pp. 15–19.

[40] Dhome, M., Richetin, M., Lapresteé, J. T., and Rives, G., *"Determination of the Attitude of 3-D Objects from a Single Perspective View,"* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 12, 1989, pp. 1265–1278.

[41] Attia, M., Slama, Y., and Kamoun, M. A., *"On Performance Evaluation of Registration Algorithms for 3-D Point Clouds,*" Proceedings of the 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV), Beni Mellal, 2016, pp. 45–50.

[42] Gold, S., Lui, C. P., and Rangarajan, A., *"New Algorithms for 2-D and 3-D Point Matching: Pose Estimation and Correspondence,*" Pattern Recognition, Vol. 31, No. 8, Aug. 1998, pp. 1019–1031.

[43] Matas, J., Chum, O., Urban, M., and Pajdla, T.*, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions,"* Image and Vision Computing, Vol. 22, No. 10, 2004, pp. 761–767.

[44] Dalal, N., and Triggs, B., *"Histograms of Oriented Gradients for Human Detection,"* Proceedings—2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, Vol. I, IEEE Publ., Piscataway, NJ, 2005, pp. 886–893.

[45] Gao, Xiao-Shan; Hou, Xiao-Rong; Tang, Jianliang; Cheng, Hang-Fei (2003*). "Complete Solution Classification for the Perspective-Three-Point Problem"*. IEEE Transactions on Pattern Analysis and Machine Intelligence. 25 (8): 930–943.

[46] Lepetit, V.; Moreno-Noguer, M.; Fua, P. (2009). *"EPnP: An Accurate O(n) Solution to the PnP Problem*". International Journal of Computer Vision. 81 (2): 155–166.

[47] Xu, C., Zhang, L., Cheng, L., andKoch, R., *"Pose Estimation from Line Correspondences: A Complete Analysis and a Series of Solutions*," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, No. 6, 2017, pp. 1209–1222.

[48] Ansar, A., and Daniilidis, K., *"Linear Pose Estimation from Points or Lines,"* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 5, 2003, pp. 578–589.

[49] Mirzaei, F. M., and Roumeliotis, S. I.*, "Globally Optimal Pose Estimation from Line-Plane Correspondences,"* 2011 IEEE International Conference on Robotics and Automation, Shanghai, 2011, pp. 5581–5588.

[50] Lowe, D. G., *"Three-Dimensional Object Recognition from Single Two-Dimensional Images,"* Artificial Intelligence, Vol. 31, No. 3, 1987, pp. 355–395.

[51] Sharma, S. , Beierle, C. , and D'Amico, S.. "*Pose Estimation for Non-Cooperative Spacecraft Rendezvous Using Convolutional Neural Networks,"* September 19, 2018.

[52] European Space Agency, "*Kelvins - ESA's Advanced Concepts Competition Website*," https://kelvins.esa.int. Accessed Januray 4, 2019.

[53] Torr, P. H. S., and A. Zisserman. "*MLESAC: A New Robust Estimator with Application to Estimating Image Geometry*." Computer Vision and Image Understanding. Volume 78, Issue 1, April 2000, pp. 138-156.

[54] Moré, J. J. "*The Levenberg-Marquardt Algorithm: Implementation and Theory."* Numerical Analysis, ed. G. A. Watson, Lecture Notes in Mathematics 630, Springer Verlag, 1977, pp. 105–116.

[55] A. Valmorbida, M. Mazzucato, S. Tronco, M. Pertile, and E. Lorenzini,"*Design of a ground-based facility to reproduce satellite relative motions*," in 2017 IEEE International Workshop on Metrology for AeroSpace (MetroAeroSpace), June 2017, pp. 468–473.

[56] **http://www.vision.caltech.edu/bouguetj/calib_doc/**

[57] Sansone, F., Branz, F., Francesconi, A. *"A Relative Navigation Sensor for Cubesats Based on Retro-reflective Markers".* 10.1109/MetroAeroSpace.2017.7999529.

[58] S. Chiodini, R. G. Reid, B. Hockman, I. A. D. Nesnas, S. Debei and M. Pavone, "Robust Visual Localization for Hopping Rovers on Small Bodies," 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, 2018, pp. 897-903.

[59] Chiodini, S., Pertile, M., Giubilato, R., Salvioli, F., Barrera, M., Franceschetti, P. & Debei, S.. *"Experimental Evaluation of a Camera Rig Extrinsic Calibration Method Based on Retro-reflective Markers Detection.* Measurement (2018). 140. 47-55. 10.1016/j.measurement.2019.03.036.