Visual Tracking and Motion Estimation for an On-orbit Servicing of a Satellite

Nassir Workicho Oumer

Dissertation
zur Erlangung des Doktorgrades der Naturwissenschaften (Dr.rer.nat.)
im Fach Informatik

eingereicht im
Fachbereich Mathematik/Informatik
der Universität Osnabrück

Osnabrück, im März 2016

Gutachter:
Prof.Dr.-Ing Peter Reinartz
Prof.Dr.-Ing Manfred Ehlers

Mitglieder der Prüfungskommission:
Prof.Dr.-Ing Peter Reinartz
Prof.Dr.-Ing Manfred Ehlers
Dr.-Ing Michael Suppa
Dr.-Ing Christine Pohl

Tag der Disputation: 29. Juli 2016

Abstract

This thesis addresses visual tracking of a non-cooperative as well as a partially cooperative satellite, to enable close-range rendezvous between a servicer and a target (client) satellite. The servicer satellite approaches the target to capture its region of interest (e.g. its nozzle) and perform orbital life extension operations such as refueling and repairing. Tracking and estimation of relative motion between the servicer and the client is a critical ability for rendezvous and proximity operation. For this purpose, LIDAR has been widely employed in cooperative rendezvous and docking missions. However, LIDAR has high weight and rotating parts, and consumes more power. On the other hand, inexpensive on-board cameras can provide an effective solution, working at a wide range of distances. However, conditions of space lighting are particularly challenging for image-based tracking algorithms, because of the direct sunlight exposure, and due to the glossy surface of the satellite that creates strong reflection and saturation which leads to difficulties in tracking procedures.

In order to address these difficulties, the relevant literature is examined in the fields of computer vision, and satellite rendezvous and docking. Two classes of problems are identified and relevant solutions, implemented on a standard computer are provided. Firstly, in the absence of a geometric model of the satellite, the thesis presents a robust feature-based method with prediction capability in case of insufficient features, relying on a point-wise motion model. Secondly, we employ a robust model-based hierarchical position tracking method to handle change of image features along a range of distances, and localize an attitude-controlled (partially cooperative) satellite. For this purpose, a histogram-based foreground and background segmentation is directly integrated into the position optimization loop in 3D space at mid-range, followed by an edge-based localization of the nozzle of the satellite at close-range. Moreover, the thesis presents a pose tracking method addressing ambiguities in edge-matching and pose detection algorithm based on appearance model learning. For the validation of the algorithms, real camera images and ground truth data are used, which have been generated with a laboratory test bed that provides illumination conditions, satellite motion and optical characteristics similar to space. Comparisons with state of the art methods are performed.

In conclusion, both motion and pose tracking approaches are extensively validated using ground hardware and software systems simulating space environment, optical characteristics and motion of a full scale satellite. The experimental results indicate that camera based methods provide robust and accurate tracking for the approach of malfunctioning satellites in spite of the difficulties associated with specularities and direct sunlight. Also exceptional lighting conditions associated to the sun angle are discussed, aimed at achieving fully reliable localization systems in a certain mission.

Zussammenfassung

Diese Doktorarbeit befasst sich mit dem visuellen Tracking eines nicht-kooperativen sowie eines teilweise kooperativen Satelliten, um Rendezvousmanöver zwischen einem Servicer- und einem Target-Satelliten (Client-Satelliten) im Nahbereich zu ermöglichen. Hierfür nähert sich der Servicer- dem Target-Satelliten um einen Zielbereich zu erfassen (z.B. dessen Düse) um Maßnahmen zur Lebensverlängerung wie z.B. Auftanken oder Reparaturen durchzuführen. Tracking und relative Bewegungsschätzung zwischen dem Servicer- und dem Client-Satelliten sind essentiell für Rendezvous- und Annäherungsoperationen. Zu diesem Zweck wurde und wird meist LIDAR in kooperativen Rendezvous-und Docking-Missionen eingesetzt. Die Nachteile von LIDAR sind unter anderem ein hohes Gewicht, rotierende Teile und der hohe Energieverbrauch des aktiven Sensors. Im Gegensatz hierzu können kostengünstige Onboard-Kameras eine effektive Lösung bieten, da sie ebenfalls bei einer Vielzahl von Aufnahmedistanzen funktionieren. Allerdings sind die Beleuchtungsbedingungen im Weltraum eine besondere Herausforderung für Algorithmen zum bildbasierten Tracking. Die direkte Sonneneinstrahlung und zahlreiche glänzende Oberflächen eines Satelliten, die starke Reflexionen und gesättigte Bildbereiche hervorrufen, führen zu Schwierigkeiten beim Tracking.

Um Lösungen hierfür zu finden, wird die relevante Literatur in den Bereichen Computer Vision und Satelliten-Rendezvous und -Docking untersucht. Zwei Klassen von Problemen werden identifiziert und für diese relevante Lösungen werden auf einem Standard-Computer implementiert. Zuerst wird in dieser Arbeite, ohne Nutzung eines geometrischen Modells des Satelliten, ein robustes merkmalbasiertes Verfahren mit Vorhersagefähigkeit bei ungenügenden Merkmalen gezeigt, welches auf ein punktweises Bewegungsmodell zurückgreift. Als Nächstes verwendet der Autor eine robuste, modellbasierte, hierarchische Positionstracking-Methode um die Veränderungen von Bildmerkmalen über einen Distanzbereich zu handhaben, damit ein teilweise kooperativer Satellit mit funktionierender Lageregelung lokalisiert werden kann. Für den mittleren Distanzbereich wird eine auf Histogrammen basierende Vordergrund- und Hintergrundsegmentierung direkt in die Optimierungsschleife für die Position, welche im 3D-Raum durchgeführt wird, integriert. Für den Nahbereich wird eine kantenbasierte Lokalisierung der Satellitendüse verwendet. Des weiteren wird in dieser Arbeit eine Methode zum Tracking der Pose gezeigt, welche auch Mehrdeutigkeiten berücksichtigt, die beim Kantenmatching und bei Erkennungsalgorithmen für die Pose mittels Erscheinungsmodell-Lernen auftreten. Für die Validierung der Algorithmen werden Kamerabilder sowie Referenzdaten aus einem Testlabor genutzt, welches Beleuchtungsbedingungen, Satellitenbewegungen und optische Eigenschaften ähnlich zu

denen im Weltraum bietet. Vergleiche mit Methoden des aktuellen Stands der Technik werden durchgeführt.

Abschließend werden sowohl Ansätze zur Bewegungs- als auch Lageschätzung umfassend validiert, indem bodengestützte Hardware- und Software-Systeme genutzt werden, welche die Weltraumumgebung, die optischen Eigenschaften und die Bewegung eines maßstabsgetreuen Satelliten simulieren. Die experimentellen Ergebnisse zeigen, dass kamerabasierte Methoden, trotz der Schwierigkeiten mit Spiegelungen und direkter Sonneneinstrahlung, robustes und genaues Tracking für die Annäherung an einen nicht funktionierenden Satelliten ermöglichen. Zudem werden in der Arbeit außergewöhnliche Lichtbedingungen aufgrund des Sonneneinstrahlwinkels behandelt um ein komplett zuverlässiges Lokalisierungssystem für eine bestimmte Mission zu erreichen.

Acknowledgment

This thesis has been written during my employment at the Robotics and Mechatronics Center of German Aerospace Center (DLR) in Oberpfaffenhofen, Germany. I have been supported while writing this thesis by many people and I am grateful for that.

First of all, I would like to thank the former and the current head of the Institute of Robotics and Mechatronics Prof. Gerd Herzinger and Prof. Alin Albu-Schaeffer, and the former Head of the Department of Perception and Cognition Dr. Michael Suppa for giving me the opportunity to work in this institute. I am deeply grateful to Dr. Michael Suppa for always encouraging and motivating me in my work. I am particularly very grateful to Prof. Peter Reinartz for supervising my Ph.D. thesis and offering me invaluable support and advice. I am grateful to Dr.habil Giorgio Panin who helped me learn important skills and discussed several topics with me. I am also grateful to the German Academic Exchange Service for the DLR-DAAD fellowship during my first three years, that led me to employment at DLR.

I would also like to thank my colleagues who helped me during writing the Ph.D. thesis. I would like to thank, Dr. Simon Kriegel and Phillip Schmidt for their invaluable feedback and Dr. Haider Ali who enlightened me during tough times. I am also grateful to Martin Lingenauber for the useful feedback. I would also like to thank Dr. Ulrich Hillenbrand and Roberto Lampariello for the valuable discussion and support. I am thankful to Dr. Klaus Strobl for helping me gain valuable hands-on experience and skills with camera calibration, and late Dr. Toralf Boge for his technical support at the EPOS facility.

Many thanks goes to my wonderful wife Shukrema for her continuous invaluable support and patience as well encouragement to do my Ph.D. I would also like to thank my son Ayan and my daughter Inaaya for sacrificing their time as well as cheering me up with their smiles.

The last, but not the least, I thank the Almighty God for giving me strength and ability to complete this work in the time He ordained.

Munich, 2016
Nassir Workicho Oumer

Contents

List of Figures

List of Tables

Abbreviations

DEOS	**DE**utsche **O**rbitale **S**ervicing Mission (German Orbital Servicing mission)
OOS	**O**n-**O**rbit **S**ervicing
RVD	**R**endez**V**ous and **D**ocking
GEO	**G**eostationary **E**arth **O**rbit
LEO	**L**ow **E**arth **O**rbit
LIDAR	**L**ight **D**etection **A**nd **R**anging
ATV	**A**utomated **T**ransfer **V**ehicle
DART	**D**emonstration of **A**utonomous **R**endezvous **T**echnology
GPS	**G**lobal **P**ositioning **S**ystem
RGPS	**R**elative **G**lobal **P**ositioning **S**ystem
RADAR	**RA**dio **D**etection **A**nd **R**anging
CAD	**C**omputer-**A**ided **D**esign
MLI	**M**ulti**L**ayer **I**nsulation
ISS	**I**nternational **S**pace **S**tation
DLR	**D**eutches Zentrum für **L**uft- und **R**aumfahrt (German Aerospace Center)
EPOS	**E**uropean **P**roximity **O**perations **S**imulator
AOCS	**A**ttitude and **O**rbit **C**ontrol **S**ystem
MAV	**M**icro **A**erial **V**ehicle
SSD	**S**um of **S**quared **D**ifferences
1D	One-**D**imensional
2D	Two-**D**imensional
3D	Three-**D**imensional
SO(3)	**S**pecial **O**rthogonal groups in three-dimension

SE(3)	**S**pecial **E**uclidean groups in three-dimension
DOF	**D**egrees **O**f **F**reedom
TIS	**T**arget **I**llumination **S**ystem
SIFT	**S**cale **I**nvariant **F**eature **T**ransform
TCP	**T**ool **C**enter **P**oint
PnP	**P**erspective-**n**-**P**oint
LSE	**L**east **S**quare **E**rror
AGE	**A**bsolute **G**eometric **E**rror
RANSAC	**RAN**dom **SA**mple **C**onsensus
EKF	**E**xtended **K**alman **F**ilter
IEKF	**I**terated **E**xtended **K**alman **F**ilter

Symbols

	Background: Lighting, Surface Reflectance and Calibration
λ	wavelength
N_p	surface normal
L	light direction
θ	angle between a surface normal and light direction
$I(x, y)$	image intensity at a pixel
(x, y)	a feature point or a pixel in an image
f	focal length of a camera
(f_x, f_y)	focal length in pixel
(c_x, c_y)	principal point of an image
K	camera calibration matrix
${}^cT_o$	transformation from a camera to an object coordinate
A_i	set of robot poses
B_i	set of camera poses
	Model-free Motion Estimation: Velocity and Prediction
$\mathbf{p}$	3D point in space
$\mathbf{u}$	2D point on image plane
$\mathbf{V}$	linear velocity of a rigid object
ω	angular velocity of a rigid object
Ω	velocity components of a rigid object
Z	depth or translation along optical axis of a camera
$\mathbf{z}_k$	measurements
$\mathbf{x}_k, \mathbf{s}_k$	states
$p(\mathbf{s}_k \vert \mathbf{z}_k)$	the probability of the state $\mathbf{s}_k$ given the measurements $\mathbf{z}_k$

A_k	transition matrix relating current and previous states
B_k	control input of a system
Δt	time between two camera frames
ρ	process noise of a system
ν	measurement noise
H_k	measurement matrix
$\mathbf{h}$	non-linear function mapping state and measurements
K_k^i	Kalman gain at iteration i
$P_{k,k-1}$	state covariance a priori
$\mathbf{x}_k^{i+1}$	state at iteration i+1
Q	process noise covariance
R	measurement noise covariance

Model-based Position Localization: Tracking and Detection

T	position vector
$f(T)$	foreground region of an image at a position hypothesis T
$b(T)$	background region
$P_f^h(T)$	histogram of foreground region at a position hypothesis
$P_b^h(T)$	histogram of background region
N_c	number of histogram cells
L_1	absolute geometric error
r_{min}	minimum radius of a circle
r_{max}	maximum radius of a circle
Z_{min}	minimum distance
Z_{max}	maximum distance
R_c	radius of a 3D circle
T_g	general rigid body transformation matrix
R	rotation matrix
$\mathbf{t}$	translation vector
T_t	rigid body transformation matrix at time t
μ	linear and angular velocities in local coordinate of T_{t-1}
G_i	Lie basis generators
$\mathbf{x}$	3D point in a homogeneous coordinate

$\mathbf{y}$	projection of a 3D point onto image plane
$\pi()$	a perspective projection of a 3D point to 2D coordinates
K	projection matrix, related to calibration matrix
σT	transformation matrix corresponding to local pose parameters μ
$\mathbf{e}_i$	re-projection error
J	Jacobian, partial derivatives of residual
w_i	weights for outlier rejection in Tukey M-estimator
W	diagonal matrix containing weights w_i
σ	standard deviation of inlier data
Ω	rotation vector
$\mathbf{v}$	linear velocity vector
ω	angular velocity vector
$\hat{\omega}$	skew-symmetric matrix corresponding to angular velocity vector
$\hat{\Omega}$	skew-symmetric matrix corresponding to rotation vector

Model-based Pose Localization: Pose Tracking and Detection

d	distance of an object from a camera
l	size of an object in an image
L_o	size of an object
$\mathbf{v}$	inter-frame motion, linear and angular velocities
Δt	time interval between two camera frames
E_k, E_{k+1}	pose at time k and k+1
$E_{k+1,k}$	predicted pose at time k+1 given pose at k
$\mathbf{x}$	the states of the filter estimate
$\boldsymbol{\rho}_{pose}$	error in pose estimation
$\boldsymbol{\rho}_v$	error in inter-frame motion estimation
$\mathbf{\Sigma}$	state error covariance
F	process model, transition matrix
$\mathbf{Q}$	state noise error covariance
S	residual covariance (innovation)
K_k	Kalman gain
E^m_{k+1}	the pose measurement at time k+1
R	measurement covariance

H	observation model, measurement matrix
B_f	number of branches of a vocabulary tree
L_f	depth or level of a vocabulary tree

Chapter 1

Introduction

Ever since satellites have been placed in orbits, the capability to access and service such precious space assets has remained a growing interest. The success of servicing the Hubble telescope and the International Space Station has been an important milestone for unmanned on-orbit servicing (OOS) technology. In particular, spectacular success has been achieved in rendezvous and docking (RVD) for the experimental mission Gemini, where Gemini 6 maneuvered to and kept in station with Gemini 7 within a few centimeter distance (Fig. 1.1). Subsequent missions further validated rendezvous and docking, followed by a successful boosting of docked spacecrafts into higher orbits. Thus, the rendezvous and docking capability demonstrated since 1960s has played a key role in today's ambition of potential on-orbit satellite servicing (see Table A.1 in Appendix A for brief history of RVD missions).

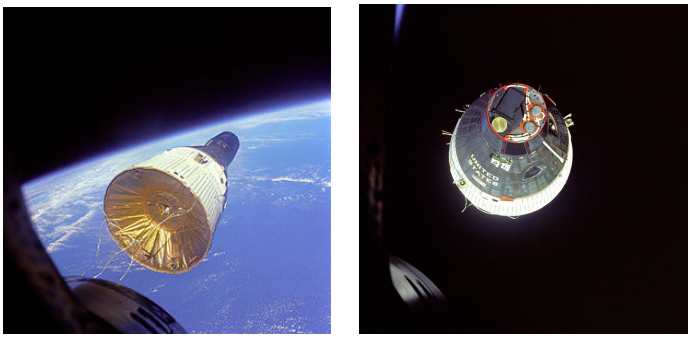


FIGURE 1.1: Rendezvous of Gemini-VII, as approached by Gemini-VI (left) and rear view of the Gemini-VII (right).

Currently, on-orbit servicing capability is crucial as the number of aging satellites has continued to increase dramatically and they occupy precious orbits, such as the geostationary Earth orbit (GEO) which is a limited resource for communication satellites. OOS requires orbital manipulation of resident objects, including on one hand the capability to remove orbital debris in order to free and reuse the orbit, and on the other hand to repair, maintain and upgrade the client satellite, thus extending its lifetime.

On-orbit servicing activities consist in an *approach* phase, where rendezvous and proximity operations are performed, and *capture* phase, which brings the servicer and target satellite in contact. In this context, the client satellite can be cooperative or non-cooperative; cooperative satellites provide features that make servicing activities easier, such as fiducial markers to help localization, or active guidance facilities to help rendezvous maneuvers. In this thesis, we consider instead non-cooperative satellites which are not pre-designed for on-orbit servicing, passive and markerless, as well as those with near end of their lifespan but still able to keep a controlled attitude (partially cooperative satellite).

In approach and capture phase, the relative position and attitude between the client and servicer must be determined. Since decades, LIDAR based measurements have been intensively employed to estimate range and pose of cooperative satellites for rendezvous and docking missions, for example videometer for Automated Transfer Vehicle (ATV), advanced video guidance sensor for DART, and rendezvous and docking sensor for ATV Jules Verne. However, LIDAR has a relatively high weight and a rotating part which wears during time, and it consumes relatively much power. In contrast, a camera is a passive and static device, with a low weight and low power consumption. The two technologies operate differently in the harsh space environment: while the LIDAR can scan texture-less surfaces and operate under relatively strong lighting, it has a limited accuracy and cannot easily take advantage of local features (such as surface edges and texture). The camera, on the other hand, relies on surface features mainly provided by the object contours, as the surface is mostly texture-less and glossy.

There exists an intense and highly directional sunlight in the space environment which results in several bright and dark shadows, i.e. easily saturates the camera sensor. On the other hand, the Earth's albedo provides a source of diffuse light that can reduce image contrast. Moreover, the surface of a satellite is usually wrapped in multilayer

insulation (MLI) material for thermal protection. Thus, the optical characteristics of the client and the space environment result in strong specular reflections, posing non-trivial difficulties for visual tracking.

This thesis aims at robust camera based tracking and motion estimation methods for rendezvous of a space object under such space lighting condition. The camera mounted on a servicer satellite is used to determine relative position and orientation (pose) of the client in the presence of strong reflection, shadow and image saturation. This work investigates and proposes an appropriate robust method to localize and estimate poses using various features of the client, for example nozzle of the satellite at close-range, sparse features, and a general satellite model depending on the structure and shape of the satellite.

1.1 Objectives

The goal of this thesis is to provide and validate robust motion estimation methods for on-orbit servicing under difficult space environment and highly reflective surface. In the present work, two possible scenarios are independently addressed; partially cooperative and non-cooperative clients. Partially cooperative targets, are temporarily functioning and under full control, but they miss a few important parts, or are about to run out of fuel. On the other hand, non-cooperative targets, instead, are full satellites or debris completely out of control, freely tumbling or spinning in space. While both targets have common optical characteristics to the space illumination, targets with tumbling motion pose further difficulty for tracking because of arbitrary motion. Regardless of the type of the target, the client motion has to be reliably estimated, tracked and predicted for on-orbit servicing mission. In this context, the overall objective of the thesis includes:

- generating realistic ground truth data and images for extensive evaluation of the state of the art and novel vision-based localization methods using RVD mock-up system [1]. The mock-up consists of terrestrial setup that reproduces satellite dynamics (such as trajectory of tumbling motion), space environment (spectrum and intensity of the Sun) and optical characteristics of the client satellite or debris.
- providing a novel and accurate motion estimation method which is relatively robust to specular reflection of the multilayer insulation (MLI) of the satellite,

- integrating robust vision based pose estimation and tracking methods particularly for close-range rendezvous and manipulation of a client satellite.

1.2 State of the Art

The success of rendezvous and docking of a satellite depends highly on relative position and orientation measurements [2]. In general, rendezvous sensors are employed to estimate relative states of satellites from far- and close-range. In many cases, ranging and GPS are used for far-range absolute navigation (up to 1 km) in Low Earth Orbit (LEO). However, GPS data may not be observable in GEO, hence ranging is mostly employed. For non-cooperative targets, a RADAR tracking from the ground provides effective solution allowing similar accuracy. For relatively lower ranges, relative navigation sensors such as RGPS, RADAR, LIDAR and Camera can be used. RADAR and LIDAR are appropriate technologies for OOS missions to reduce the relative range between the servicer and the client satellites to orders of 10 m [3].

However, the high power consumption and relatively heavy mass and volume of these sensors constrain the power and mass budget of the servicing satellite. Consequently, a camera based navigation is preferable particularly for close-range relative navigation because of its lower mass, volume and power consumption. The camera based navigation, however, depends on the distinguishable image features of the client satellite, which are limited within a relative distance of several tens of meters between servicer and client. For this reason, the state of the art rendezvous of a satellite relies on the LIDAR and cameras with retro-reflectors on the client. However, the client satellite in orbit servicing missions is assumed either non-cooperative or partially cooperative, hence not accessible for mounting visual aids for camera based navigation.

On the other hand, camera based markerless 3D motion estimation has been widely practiced in robotics and industrial application. For this purpose, model based methods [4, 5] which, match re-projected model edges to observed edges in the image, have been used for tracking of texture-less objects. Stereo camera based tracking that exploits optical flow methods [6] and descriptor matching [7] serves the same purpose in the absence of geometric model. These methods provide desirable accuracy and robustness under normal illumination. However, as stated earlier, illumination condition in space

FIGURE 1.2: DEOS: German on-orbit satellite servicing robotic mission for rendezvous, capture and docking of client satellite. After the servicer (right) approaches the client (left), a robot arm mounted on the servicer grasps the client and performs the desired operations.

and the reflective property of the surface of the satellite hinders to exploit the state of the art vision based methods for 3D motion estimation.

This thesis, inspired by DEOS on-orbit servicing mission (Fig. 1.2), aims to mitigate the problem of camera based 3D motion estimation and tracking due to illumination and reflective surface properties using an utmost realistic test environment.

1.3 Thesis Outline

This work addresses motion estimation and tracking methods, for on-orbit satellite servicing. The targets of the visual tracking include malfunctioned satellites or debris in the absence of their geometric models and partially or fully non-cooperative satellites in the presence of geometric models. Motion estimation and tracking hereafter refers to the visual tracking in the absence of a computer-aided design (CAD) model, while pose tracking refers to visual tracking in the presence of the CAD model of the satellite.

1.3.1 Contributions

The thesis consists in novel contributions to the development of vision-based localization for on-orbit servicing. It focuses on realistic images similar to space scenario as a part of applied research in the challenging space environment.

From an experimental point of view, this thesis provides extensive ground truth data and real camera image sequences for evaluation of vision-based localization algorithms

for rendezvous and capture of a satellite based on a ground hardware system, consisting of satellite motion, its optical characteristics and space environment. The experimental data which consists of image sequences and associated ground truth motion trajectories, is built so as to reflect various satellite poses including spinning and tumbling, specular reflections and shadows. Moreover, to account for hardware constraint for providing ideally noiseless ground truth and accurate spectrum of the Sun, photo-realistic rendering is employed.

From the methodological point of view, the thesis on the other hand, explores model-free feature-based tracking and model-based localization methods and makes advancements in the regard to robustness and accuracy. Moreover, we follow the current space standard and rely on a standard computer with a single thread implementation for real-time application.

Firstly, we present a novel model-free motion tracking method in 6 degrees of freedom, with predictive capability in the absence of sufficient features. Previous model-free approaches such as [8] attempt to recover relative pose, from set of 3D points reconstructed with stereo and register [9–11] them, or from structure and motion estimation (for example, [12]). The former, is well known in ego-motion estimation and provides reasonable accuracy and robustness for relatively low noise camera images and objects rich in texture, while the later is highly prone to local minimum due to large free parameters during minimization. The present work, in contrast, takes account of relatively less textured satellite and employ stereo tracking approach, by integrating point-wise kinematics model of each feature into Bayesian filter such as Kalman filter to help predict the motion, also in case of any missing measurements.

Secondly, we provide a robust model-based hierarchical position localization method to handle change of image features along a range of distances, and localize a partially cooperative satellite. A histogram-based foreground and background segmentation is directly integrated into the position optimization loop in 3D space at mid-range, followed by an edge-based localization of the thrust engine (nozzle) of the satellite at close-range. A robust model-based nozzle detection method, relying on stereo images is also provided. Moreover, we adapt the state of the art model and image edge alignment methods to handle strong specular reflection and shadow for pose tracking of fully non-cooperative and tumbling satellite. The proposed method adapts an edge-based tracking method [4]

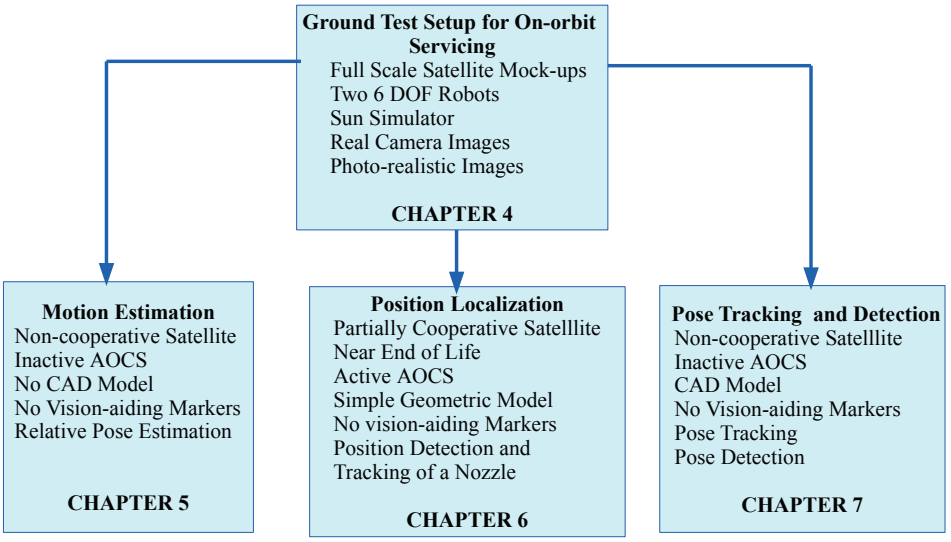


FIGURE 1.3: Graphical illustration of the main Chapters of the thesis.

by introducing adaptive search length for edge matching, and handling specularity. The related model-based satellite tracking approach [13] also employed [4] to demonstrate the performance with a scaled satellite mock-up. However, the mock-up consists of several contours (edges) and little multilayer insulation, which is relatively easier for model-based tracking. Similarly, during writing this thesis, [14] presented a satellite tracking method combining edges [15] and color statistics [16], however relying on GPU for real-time performance.

Thirdly, the thesis presents a global pose detection method, based on appearance learning which is not in literature to the best of our knowledge. The features of a mock-up satellite are used to learn the appearance under several illumination direction and views. We employ the concept of bag-of-words realized in a vocabulary tree for fast retrieval of observed image, to estimate the 3D pose from the 3D-2D feature correspondences. The 3D points are computed off-line by re-projecting training features stored in database, using Z-buffer rendering of the CAD model. The summary of the Chapters of the main contributions is presented in Fig. 1.3.

1.3.2 Overview

Literature that addresses 3D motion estimation and tracking in the context of a satellite rendezvous, docking and proximity operations are presented in Chapter 2. Moreover, the chapter explores the methods commonly used for tracking objects under normal illumination, and points the current camera based methods proposed for on-orbit satellite servicing. Chapter 2 is intended to provide general review of related literature, however the most related literature to this thesis may be repeated in the corresponding chapter.

Some background explanations are provided in Chapter 3 in order to introduce readers from various background such as aerospace engineering. This Chapter also offers a basic theoretical framework for the subsequent chapters.

Realistic data to validate and verify the methods is emphasized in the thesis. Chapter 4 details how the space environment, satellite motion and optical characteristics of its surface are terrestrially simulated using both hardware and software. The final data consists of ground truth trajectory in 6 DOF and calibrated image sequences under various illumination and motion.

The fragmented parts of a satellite (debris) need to be de-orbited in order to re-use the important satellite orbits such as GEO. Mostly such satellite debris may not have geometric model. In such cases, we have to address the 3D motion estimation problem solely based on image features. The standard method includes stereo based tracking relying on some textures. However such methods may fail in the absence of sufficient textural features. To cope with such difficulty, point-wise kinematics for feature prediction in the frame work of stereo based tracking is introduced in Chapter 5.

Despite the challenges that could be faced due to illumination and reflectance properties of satellite surfaces, the motion estimation and tracking problem is relatively relaxed in the presence of simple model (Chapter 6) or CAD model (Chapter 7). In Chapter 6, the thesis addresses position localization of a partially cooperative satellite (active attitude and orbit control system), whose attitude measurement sensors can be used to align its attitude to that of the servicer satellite. Moreover, full 6 DOF pose tracking of a tumbling satellite under various illumination is thoroughly addressed in Chapter 7. Global pose detection method is also discussed. Finally the thesis summarizes the main ideas in Chapter 8, concludes and points out the future perspectives.

Chapter 2

Related Work

The development of an accurate and robust 3D tracking system for space application such as an approach of a satellite for on-orbit servicing, requires a good understanding of the space environment, optical and dynamic characteristics of the object. The knowledge in the fields of computer vision, optimization and aerospace are vital for design and implementation of a camera based navigation system. In the context of this work, the overall objective is to provide a robust camera based localization methods for the approach phase of the on-orbit servicing. The localization succeeds either by estimating the incremental motion of the target or the current object absolute pose. These choices depend on the available geometric information of the client satellite.

The full localization system incorporates detection, (re)initialization and tracking of the object. In general detection is a global method mostly achieved by brute force approaches to search and locate the desired object of interest. However, this thesis focuses mainly on the problem of localization generally subsumed to:

- 3D motion tracking,
- 3D pose tracking.

The 3D motion tracking addresses the challenges of estimating relative position and orientation in time from image data in the absence of geometric information of the client satellite. This involves generally image processing, camera calibration and high-level interpretation of the image to extract valuable geometric information. On the

other hand, 3D pose tracking serves the same purpose while exploiting the geometric information of the client satellite to estimate the absolute pose. The model of the satellite in general provides a priori information to constrain the pose estimation process by reducing the search space.

The following sections review the relevant scientific literature in the above described areas. In particular section 2.1 reviews the technology mostly used in ranging and state estimation during cooperative rendezvous and docking missions, highlighting potential problems as applied to non-cooperative systems. The general visual tracking problem in computer vision is addressed in section 2.2 by classifying the approach into feature-based and model-based. Each approach is thoroughly reviewed with respect to on-orbit servicing under space lighting condition.

2.1 Rendezvous and Capture of a Satellite

The major idea of on-orbit servicing of a tumbling satellite is initiated since space debris (see A.2 in Appendix A) has been concern for satellites. The successful satellite rendezvous and docking capability has also built strong confidence for possible realization. For example, the LIDAR based measurements have been technologically proved for estimating range and pose of a cooperative satellite. To name a few specific missions, videometer was used for rendezvous mission of the Automated Transfer Vehicle (ATV), advanced video guidance sensor (AVGS) for DART mission, rendezvous and docking sensor for the mission of ATV Jules Verne, and LIDAR based autonomous navigation of Hayabusa for rendezvous of asteroid Itokwa (Hayabusa is a remote sensing mission, a lander and a sample return). Most recently PRISMA in-orbit test bed [17] was launched to demonstrate technologies for formation flying, rendezvous and on-orbit servicing capabilities. A vision based sensor (VBS), in addition to GPS based relative and absolute navigation system and radio frequency (RF), was employed for various maneuvers and approach of Mango (servicer) and Tango (client) satellites of PRISMA.

LIDAR based navigation for rendezvous and docking of a satellite in several missions is, thus a well established method. However, as stated earlier, the constraint of limited power and mass budget as well as the smaller field of view (particularly at close range) motivates to look for better solutions such as vision based methods [18–20].

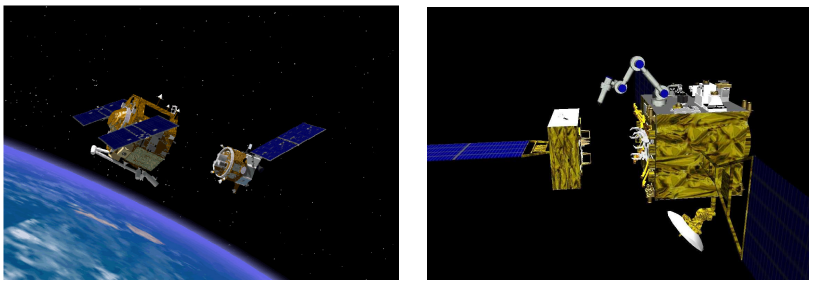


FIGURE 2.1: On-orbit servicing missions: ASTRO and NEXTsat satellites of Orbital Express (left), and Hikoboshi and Orihime satellites of ETS-VII (right).

In a few robotic satellite missions such as ETS-VII (Fig. 2.1), camera based navigation for proximity operations and docking [18] is technologically proved. In this mission, optical markers were mounted on the client satellite to aid 6 DOF pose tracking for visual servoing. These markers were well localized in the known model of the client, hence can be easily extracted in the image. On the other hand, SMART-OLEV [20] on-orbit servicing system conceptualized stereo camera based navigation (far-, mid-, and close-range). Particularly a vision system for guidance and approach from 5 m consists of a stereo camera target illumination system to illuminate the client when in shadow or within the apogee engine nozzle.

The experimental mission of Orbital Express [19] conducted by DARPA demonstrated successful rendezvous and capture experiments using a servicer (ASTRO) and target (NextSat) satellites (Fig. 2.1). NextSat was pre-designed for on-orbit servicing, which facilitates relative navigation by providing essential features for tracking. The vis-STAR tracking software [21], which integrates various rendezvous and capture sensors such as laser range finder, visible cameras, infrared cameras along with laser-based advanced video guidance sensor, provided primary relative navigation at mid- to long ranges. The vis-STAR software is a real-time image correlation algorithm which exploits a priori knowledge about the client, enabling a relative navigation to a passive client. At short range (60 m approach to capture) rendezvous sensor data was combined with AVGS data to provide fully redundant measurements of bearing, range and relative attitude. In addition to the far- and close-range approaches, the camera is used by a robot arm for vision-feedback to capture the client (NextSat).

The successful demonstrations made so far by employing redundant vision based navigation for formation flying and on-orbit servicing, have indicated that a camera based navigation can substitute laser and radio frequency based rendezvous sensor methods.

These successful missions however rely on either navigation aiding markers or clients predesigned for an on-orbit servicing . Therefore, a vision based tracking methods should be experimentally proved in general on non-cooperative clients, not designed for on-orbit servicing and without vision aiding markers. The next section presents review of computer vision methods for motion estimation and tracking, and addresses the challenges as they are applied to mostly specular objects such as satellites.

2.2 Motion Estimation and Tracking

Camera based tracking extracts 3D motion or pose of an object from sequence of images. In the context of ego-motion, which refers to the motion of the camera, 6 degrees of freedom (DOF) motion of the object needs to be estimated to localize the object relative to the environment. Similarly, when the object of interest for tracking is the moving target, the 6 DOF motion is estimated relative to either a fixed camera frame or another moving object. Despite the difference of what is to be tracked (an external target or self), the approach of the tracking shares a similar methodology. There exist several established tracking methods developed by the computer vision community under nominal illumination conditions. Based on the available a priori information of the object, these methods can be broadly classified into model-free and model-based methods.

2.2.1 Model-free Motion Tracking

Model-free 3D tracking methods utilize 2D image information to recover motion of the object. In the context of robot navigation using a monocular camera, SLAM (real-time structure from motion) methods [22–24] simultaneously estimate structure and motion, while other methods such as [25] split the tracking (motion) and mapping (structure) into separate tasks processed in parallel. On the other hand, fully dense tracking and mapping [26] has been employed relying on general purpose GPU for real-time camera motion estimation. More recently a semi-dense monocular visual SLAM and odometry techniques [27, 28] use intensity values in the image.

Monocular SLAM approaches are widely used for weight and power critical applications such as micro aerial vehicles (MAV). Such methods cannot provide the scale of the estimated structure due to monocular view. Consequently, stereo based approaches [29–31]

are employed in robot visual odometry. The multiview approach benefits in two ways; it reduces the number of free parameters to be estimated, and provides more data in wide base line for geometric interpretation and is less sensitive to noise than monocular approach, despite the difficulty of finding correspondences. In order to improve matching accuracy and robustness more than two views are employed [32, 33]. In [32], features are detected in each incoming frame and matched using normalized sum of squared differences (SSD). The trinocular cameras, mounted two horizontally and one vertically above, provide improved matching accuracy. On the other hand, a quadrifocal method [33] which uses a stereo camera but performs dense spatial matching between temporal stereo image pairs. Inspired by [33], ego-motion estimation [34] based on Lie algebra transformation representation, exploits bilateral symmetric transfer function to propagate the last 3D set of points to the current with relative motion parameters, while the structure is estimated from stereo cameras.

Dense SLAM and odometry methods are generally believed to provide an accurate and a robust motion and structure estimation in spite of computational inefficiency. However, in the context of motion and structure, [35] argues for feature-based methods. This is because of photometric and geometric invariance, optimal estimation due to uncorrelated errors between features and computational efficiency and convergence. Moreover, feature-based methods are natural choice when the imaged object loses its dense structure due to specular reflection, sensor saturation and shadow. Therefore, this thesis focuses on feature-based methods of motion estimation.

Feature-based approaches focus on tracking 2D features such as geometric primitives (segments, circles), distinct point features, object contours and regions of interest. The primary motivation for adapting feature-based method is due to its geometric and photometric invariance and real-time capability on standard processors. Point features are relatively easy for 3D tracking and pose estimation in the absence of an object model. For this purpose, there exist several point feature extraction approaches, which are good in localization accuracy [6, 36, 37], computational efficiency [38], scale invariance [7, 39, 40] and recently developed methods based on scale invariance for efficient matching and tracking [41, 42] as well as for memory critical applications [43]. However, the performance of each detector and descriptor depends on the context of application such as recognition and tracking as well as content of textural information of the object or the scene.

The standard approach of stereo motion estimation follows a common structure as employed in [8]; first image points are detected and matched across stereo pairs, which results in disparity image. The disparity image is used for reconstruction in Euclidean space. Then the rigid motion estimation is performed based on 3D correspondences using SVD [44], quaternion [45], quaternion supported with bootstrap [11], dual-number quaternion [9] and based on least squares minimization of two 3D point sets [10]. These 3D point registration approaches assume the reconstruction error is independent, and minimize the least square error incurred during registration to estimate 6D Motion parameters. On the other hand, [46] derived motion estimation method in disparity space which leads to optimal minimization of disparity and feature detection errors based on the fact that unlike Euclidean space, noise in disparity image is homogeneous and isotropic. Also a hybrid stereo tracking approach [47] exploited ICP for registration of the 3D point sets and normal flow to constrain the motion to handle noisy depths, and small and large movements.

Rendezvous and Docking Applications

Despite several of the camera based methods (real-time monocular SLAM, visual odometry, and stereo-based methods) are employed for robot navigation under normal lighting, there exist a few related articles [12, 48, 49] which employ similar techniques for satellite rendezvous and docking. The stereo based method in [48] provided 3D measurement and motion tracking in the absence of geometric model by considering two cases. In the first case, a reference frame is built as soon as Euclidean coordinates of 3 points in camera frame are available. The origin of this reference frame is the first point, and the second point defines the X-axis and the third point defines the X-Y plan. The motion description is then provided by the centroid of these points, and quaternion associated with SO(3) map related to this built reference frame. Secondly, the stereo camera system itself builds the possible object model in the camera reference frame. The model is defined, by using target measurements during the time interval in which the object is kept still. However, this method has drawbacks; the 3 point features may not be visible and tracking with the minimum three features is not robust to change of lighting and noise. Furthermore, the built model may not be accurate enough due to specular reflection. A better technique, instead, employed a stereo system aided with a pattern projector [50] to estimate the structure of the satellite for modeling. The pattern projection is used to

tackle reconstruction problems due to specular reflection in spite of introducing active system.

On the other hand, particle filter based structure and motion estimation method [49] provided a frame-to-frame tracking of a tumbling object, relying on GPU for real-time efficiency and SIFT features. A hybrid method of prediction is introduced to tackle the problem arising due to the unknown center of mass of the target. The Rao-Blackwellised Particle filter uses the process model for prediction of rotation and translation is predicted through measurement inversion to avoid large covariance of process model. However, the experiments did not indicate any validation on satellite images or simulation. Related to SFM, [12] employs a stereo camera (without stereo matching) to recover pose, velocities and structure as well as inertial parameters of a non-cooperative satellite using EKF filtering scheme. In this scheme, inertial parameters are identified through a hypothesis-based likelihood score, over a finite set of possibilities, and numerical simulation have been presented, assuming exactly 10 feature points of which the locations are uniformly spread over the target surface. The main disadvantage of this method is estimating several parameters in large search, which can often fall in to local minima.

2.2.2 Model-based Pose Tracking

The robustness and accuracy of visual tracking can be enhanced by incorporating a priori information of the object. Model-based methods rely on geometric information (CAD model) of the object for robust tracking. The knowledge about the scene provides implicit 3D information which allows improvement of performance by enabling to predict hidden movement of the object and reduce effects of outlier during tracking.

There exist several 3D model-based tracking methods which arise mostly due to the employed:

- feature type, e.g point, edge, texture,
- pose parameter parametrization,
- registration method.

Various images features have been used in literature for model-based tracking, including point features [51–53] and contours [4, 54, 55]. Moreover segments, conical and

cylindrical objects [56–58] have been used for 3D pose tracking and estimation. While primitive features such as conics and cylinders are rarely available in the object, most human-made objects consist of contours and internal edges, which is widely used for model-based tracking. Furthermore, image edges are robust to illumination and view variation, and therefore this thesis focuses on contour (edges, lines) based tracking in particular for texture-less objects with a given 3D model.

Early edge-based pose tracking traces back to the so called RAPiD tracker [59], where the 3D object model consists of high contrast edges such as surface markings, folds and profile edges. RAPiD uses only a few selected control points along the edge. At these control points, the search region for the edge is normal to the edge. The pose parameters are computed by linearizing with respect to image motion. Later [4] introduced Lie algebra to represent motion [60] since the group SE(3) exactly represents the space of poses that form the output of the system. The Lie algebra is the tangent space to the group at the identity and, is therefore the natural space to represent differential quantities such as velocities and small motions. Related to this method [15] introduced a visual servoing concept to pose estimation based on virtual visual servoing framework [56] through full scale non-linear optimization, using numerical iterative algorithm such as Gauss-Newton and Levenberg Marquadt.

Various efforts have been made to improve the robustness and accuracy of model-based tracking by employing direct methods [61], currently based on mutual information [62, 63] by utilizing all image-pixels directly in the minimization process. Such methods mainly serve for the purpose of textured object tracking under illumination changes. On one hand, to improve the pose priors for edge-based tracking, there have been various attempts to enhance the pose accuracy by incorporating: textural information and interest points [64–66], optical flow [67–69], lines with keypoint features [70], dense and sparse methods [71], model-based and model-free methods [72]. On the other hand, multiple hypothesis approaches [64, 73, 74] and [75] have been developed for robust tracking to background misleading information such as clutters.

Similarly, in order to make the tracker robust to motion blur and occlusion, region based methods [76–78] and [79, 80] as well as contour based methods that integrate color statistics around edge control points [81, 82] have been proposed.

Rendezvous and docking application

CAD model-based tracking is one of the most studied area for industrial applications. However, unlike the fully controlled environment in industrial application such as bin picking and welding, the space environment as well as optical characteristics of the satellite for the on-orbit servicing is more difficult. In order to address this problem, several attempts have been made. In literature, the solution to this problem has been approached by:

- reconstructing the 3D structure of the satellite with stereo camera and registering with its CAD model [83, 84] using iterative ICP algorithm based on [85],
- registering 2D image directly to a generic 3D model [13, 14, 86] and 2D prototype model [87], to special primitives such as apogee engine nozzle and Launch vehicle interface ring [88], partially or fully visible rectangular shapes [89], and line model and segments [90, 91] of the satellite,
- estimating inertial parameters, i.e. relative magnitude of the inertia matrix and center of mass along with the position, orientation, velocity [92] from range images.

These approaches rely heavily on structures of a satellite, paying less attention to the main problem of specular reflection associated with the multilayer insulation. An exception to these methods is [84] which attempts to reconstruct the 3D points and match with the CAD model. However, an edge-based tracking method is a better approach in the presence of the CAD model of the satellite [13, 14].

2.3 Handling of Illumination and Specularity

A practical approach used to deal with illumination variation is usually to impose certain assumptions on the scene or object so that the motion estimation algorithm [93–102] handles perturbations. The assumptions address surface reflectivity, by decoupling irradiance variation from viewer position. In particular, the scene or the object is assumed as a perfectly Lambertian surface which is usually a good working approximation. Motion estimation, strictly speaking however, depends on available distinctive features such

as corners and edges in case of feature-based tracking, or texture of the object for template tracking. In spite of plenty of such features, accuracy of motion estimation also depends on the pose of the object. The perceived image, however, depends on illumination, surface reflectivity and view direction. Typically, visual tracking methods assume nominal illumination condition, which is not always the case in practice. However, in literature several strategies have been devised to handle mild illumination variations. In general, motion estimation of Lambertian and specular objects under illumination variation can be handled differently. The illumination change for a Lambertian object can be addressed [103] by:

- robust estimation of motion,
- joint estimation of motion and illumination,
- estimation of motion invariant to illumination.

2.3.1 Robust Motion Estimation

This approach handles implicitly the effect of illumination, as generally employed to reduce the effect of perturbations at all image processing levels. Hence, estimation is not completely insensitive to the change in illumination. In particular, robust estimators are employed to suppress unwanted signals from various sources and processes such as, inconsistency between assumptions of scene reconstruction and the reality, noise during image acquisition, discrepancy between real image formation and the assumed image formation. In the context of robust estimation on feature-based visual tracking, the perturbations are mainly attributed to the position of extracted features. The accuracy of the extracted feature depends on illumination. Thus, the robust estimation implicitly handles to some extent the effect of illumination. Well known approaches are based on RANSAC [104], which search a subset of features that is consistent with the model. On the other hand, M-estimator methods are highly popular, and tend to reject measurements that are not within the common variance [4, 5, 15, 64, 105–107]. In contrast, descriptor-based motion estimation methods achieve robustness with histograms as appearance descriptors by varying the bin size according to the expected perturbations. On the other hand, appearance-based motion estimation methods pay special attention to the matching criteria between the patches, for example normalized

cross correlation [64], mutual information [108–110] and local intensity ordering criteria [104, 111] to compensate for illumination effects. Statistically robust methods such as M-estimators [112] are applied for appearance-based methods [113, 114].

2.3.2 Estimation of Motion and Illumination

This method explicitly models the acquired information to be dependent on both motion and illumination. The effect of illumination are compensated by computing the free parameters of illumination simultaneously to the parameters of motion. A simple method, however limited to planar patches and moderate illumination change, exploits global brightness and contrast [109, 115] of the image. In contrast, surface reflectivity and illumination change can be taken into account resulting in a complex model. In practice however, such models end up with only a reduced set of parameters that can be estimated [116]. On the other hand, illumination subspace method shows popularity for motion estimation in the image plane [117–119] and in three-dimensional space [61, 113, 116, 119–126].

2.3.3 Estimation Invariant to Illumination

Image features such as keypoints and edges are to some degree illumination invariant. Appearance invariance is usually obtained by separating the method from illumination dependent pre-processing steps using such features. Appearance-descriptor methods based on contour or area moments [127–129] rely on a correct image segmentation. This segmentation is a difficult task for inhomogeneous objects. Accordingly feature extraction is expected to be in large extent invariant to illumination changes. Point features such as used in [119, 130, 131] and contour or edge features as employed in the methods [81, 98, 132–134] as well as combination of these features as employed in [64, 66, 72, 135] are invariant to illumination changes, however the detection of these features is usually variant to illumination effects. On the other hand, motion estimation based on appearance is generally variant to illumination. As the object or camera moves, the irradiance can vary over time. In particular, radiance perceived from surface of non-Lambertian objects depends on the view-point due to specular reflection. Thus, there exist no method in general that is completely illumination invariant.

2.3.4 Pose Estimation under Specular Reflection

In the context of pose estimation [136–138] and surface reconstruction [139–143], there exist several contributions that deal with specular objects. A specular geometry describes that image features exist as either real or virtual; real features are directly used by vision algorithms such as matching, tracking and structure from motion. On the other hand, virtual features (specular cues), which are specular reflections of a scene or an object due to change of viewpoint, create ambiguity for visual interpretation [140].

Various methods have been proposed to utilize these virtual features for surface reconstruction and pose estimation. For example, shape and reflectance parameters are simultaneously estimated from multiple views of an object made of single material with known lighting [139]. Similarly, Phong reflectance model is used to compute reflectance of the object shape modeled with triangular mesh, and minimize non-linear least square cost function over the shape and reflectance parameters. In contrast, specular reflections produce 2D image motion (specular flow) [140] in dynamic objects. This specular flow is related to the 3D structure of textured object or scene and parametric mixture models are used to recover a surface [141]. Shape from specular reflection usually assumes a limited number of surfaces, in which its structure is known or sparse, and the environment is calibrated. In contrast, [142] presented a reconstruction approach which considers general surfaces under unknown real world environments. They recovered 3D shape from optical flow induced by relative motion between a specular object, an observer and their environment. Similarly, [143] presented variational optical flow technique, which accounts for characteristics of specularity including parabolic singularities related to surface curvature.

On the other hand, tracking an object with specular highlights [144] exploits Phong's model, by approximating general photometric changes with a continuous and differentiable function, represented with first order Taylor approximation at a pixel point in the neighborhood. This approach extends sparse optical flow tracking method [6], compensating for illumination changes and specular highlights. In the context of pose estimation, unlike classical methods which discard lighting information, [136] refined coarse pose estimates by incorporating lighting information in texture and specular cues to improve accuracy of standard template matching algorithm. In this approach an environment map is retrieved from the specular pixels of shiny objects and registration is

performed in both image and lighting environment space. Contrastingly, [137] utilized specular highlights that appear on the shiny object, which are easy to detect and robust to changes in background, texture variation and occlusion of non-highlighted parts, as cues for pose estimation.

More practical pose estimation method [138], exploited environment map, specular reflections and specular flow, background clutter and inter-reflections. However, it requires to use a calibration object such as a mirror sphere in the target scene to capture the environment map. Thus, it is inappropriate for human inaccessible environments such as satellite orbits. Despite the existence of certain reflectance models of specular objects, they are in general rarely exploited for real world problems, mainly due to their complexity and underlying assumptions such as placement of a calibration object in the scene.

2.4 Summary

The approach and manipulation of a malfunctioning satellite relies on measurements of rendezvous sensors such as LIDAR. The LIDAR based rendezvous and docking is very matured technology frequently used in several missions. However, it is well known that a LIDAR is an expensive device in terms of power and mass budget in satellite design. Hence, an attempt to exploit inexpensive camera for rendezvous of a cooperative satellite has recently increased. The visual markers on the cooperative satellite aid for an accurate and a robust vision-based tracking. In contrast, a camera based tracking of a marker-less non-cooperative satellite is yet challenging due to harsh space illumination and reflective surface of the satellite.

On the other hand, camera based motion estimation methods are widely used for applications such as robot navigation and industrial bin picking and placing. For this purpose, a model-free methods such as multiview camera (e.g stereo camera) or structure from motion, and CAD model-based methods are well known to operate under nominal illumination conditions. Several attempts have been made to improve the tracking methods by handling illumination variation and specular reflection, and introducing robust estimators into the optimization. Despite those efforts to improve vision-based methods, it is still far from direct application for space due to harsh illumination condition.

Recently a number of works have been introduced to vision-based tracking of a non-cooperative satellite. On one hand, stereo camera based motion estimation is introduced to simultaneously estimate inertial parameters of a non-cooperative satellite. Such approach risks to failure because of the need to estimate several parameters (rotation, translation, mass center, inertial matrix) and also lacks practical experimental validation that considers sunlight illumination and reflective multilayer insulation of the satellite. On the other hand, a model-based tracking of a non-cooperative satellite that matches projected CAD model lines to gradients of image (represented by line segments) is introduced. This model-based approach exploits internal and silhouette edges, and is shown to be robust to nominal space illumination, but relies heavily on GPU for processing of the CAD model. The considered non-cooperative target is either scaled or at relatively far range, thus is fully in the field of view of the camera. This property favors model-based edge tracking because of the available several features. Most importantly, the approach fails to demonstrate the influence of reflective multilayer insulation (MLI) which is difficult particularly when tracking region of interest at close-range, and limited to single illumination condition. We remark also here that graphics processing currently is not in the space software standard due to safety issues.

Therefore, this thesis presents a real-time motion estimation and tracking of a satellite with dominant MLI at close range, which is implemented on a standard processor with realistic test environment. Computer vision methods depend on the features of the object to be tracked, hence a general method of a vision-based tracking of a satellite appears to be infeasible in light of a space application. Therefore, in this thesis three conditions of the client satellite are considered: in the absence of the geometric model, in the presence of geometric model of attitude controlled and a freely tumbling satellite. In the next chapters, we provide a new stereo camera based method of motion estimation of a model-less satellite that relies on strong keypoints and provide prediction capability in the absence of missing features (chapter 5). Then we introduce a robust model-based localization methods (chapter 6 and 7) validated with realistic test bed (chapter 4) under various illumination direction.

Chapter 3

Background

This chapter introduces to image processing and 3D localization which help readers from different background better understand subsequent chapters of the thesis. We use often image processing techniques to extract unique features that can be efficiently used for localization. In 3D localization, we estimate spatial position and orientation (pose) of an object; this includes detection of an object such as a satellite in global space and frame to frame tracking in time. A camera-based pose estimation depends on several factors such as lighting, surface and geometry of the object, camera and lens system, camera projection model and calibration as well as image features. In the following sections we discuss these factors which affect a 3D localization.

In Section 3.1 we present lighting and reflection models. Section 3.2 describes principles of image sensors, image formation and the widely used pinhole camera model. A camera calibration is discussed in Section 3.3. Section 3.4 presents image features often used for localization. Pose estimation and geometric models of an object are presented in Section 3.5 and 3.6 respectively. Finally, we briefly summarize the chapter in Section 3.7.

3.1 Lighting and Reflectance Model

Lighting and material reflectance as well as geometry of an object determine intensity and color of an image. The visual perception of light depends on wavelength; a spectral density is often used to refer the dependence of light density on wave length λ, which is typically between 400 nm to 700 nm. Related to lighting and surface reflectance are

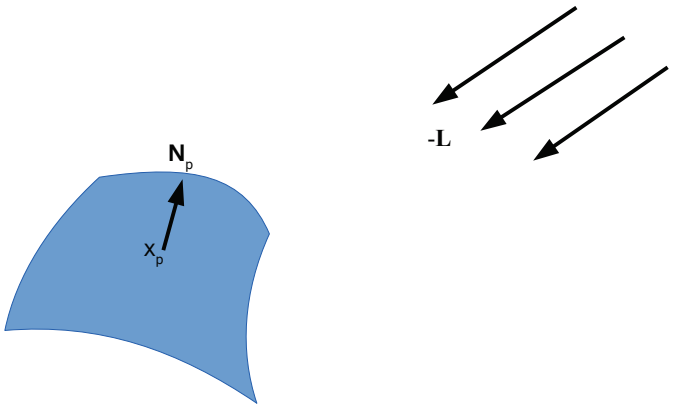


FIGURE 3.1: Distant source of light and surface of an object. The rays of a distant source of light such as the Sun are approximately parallel.

irradiance and radiance. Irradiance is a measure of light arriving at a surface per unit surface area and is central to camera sensor measurements, while radiance describes the light emitted or reflected from a surface as a function of direction.

3.1.1 Distant Light Source

For a light source that is far away such as the Sun, in unit direction L with small angular extent, the light rays are approximately parallel (Fig. 3.1). A surface patch with a surface normal N_p in the same direction of light maximizes the amount of light on it.

For a small patch of area dA_p, irradiance is the power that passes through the area of the patch projected onto a plane perpendicular to L (Fig. 3.2), i.e. $dA_L = dA_p|cos\theta| = |N_p.L|dA_p$. An opaque surfaces will be in shadow for $N_p.L < 0$.

Fig. 3.1 and 3.2 illustrate the configurations that are encountered also for on-orbit servicing scenarios, where the direction of the sun and the surface normals of the client satellite affect the camera-based localization. This is because the amount of light reflected back to the camera sensor depends on the direction of the sun and the orientation of the satellite. As the sun direction and normal of the surface of satellite coincide, much light is reflected to the camera sensor, and may result in over-saturation of the imaging sensor. The over-saturation is one of the difficult problems to address in camera-based

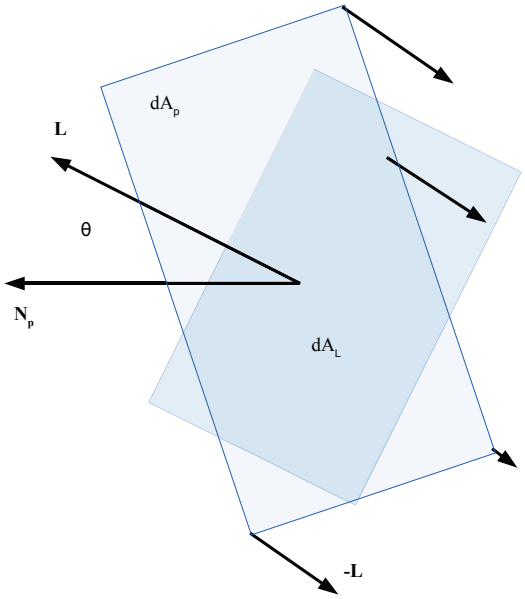


FIGURE 3.2: The angle of incidence of a distant source of light on a tilted surface of an object. A surface receives maximum light when the surface normal and the light direction are parallel.

tracking. On the other hand, the effect of shadow is higher when the angle between the surface normal and sun direction is close to 90°. In contrast to over-saturation, under-illumination can be ameliorated by an additional artificial source of light. In this case, irradiance from different light sources at a camera sensor is simply integrated. For example, in addition to the sun light, a target illuminant mounted on a servicer satellite contributes irradiance to a visual tracking camera. This integrated irradiance increases the contrast of the image of a satellite in shadow.

In addition to the geometrical configuration of the object with respect to the direction of the source of light, the reflective property of an object or its material plays a crucial role. A Bidirectional Reflection Distribution Function (BRDF) is used to model a wide range of material reflectance properties. It captures the dependence of reflectance on the directions of incidence and reflection. The BRDF is a basic radiometric concept used in computer graphics for rendering of photo-realistic images from a 3D model of a scene or an object. The BRDF can be measured from real objects using a camera and a light source. There exist, however several analytical and phenomenological BRDF models, for example Lambertian (for perfectly diffuse surface), Phong (for combination

of diffuse and specular surface) and Torrance–Sparrow model (for isotropic materials). From a theoretical perspective of the BRDF, the reflectance of a surface can be classified as diffuse and specular.

3.1.2 Diffuse and Specular Reflection

In diffuse reflectance, light is absorbed and re-emitted from the object, scattering in all directions. The pigmentation of the object determines the distribution of the reflected light. Diffusely reflecting objects appear similar from all viewing directions, i.e. radiance is independent of viewing direction. The Lambertian model is used to represent perfectly diffuse surfaces by constant BRDF.

In contrast, in specular reflection, the spectral distribution of the reflected light is the same as that of incident light. The specular reflection creates glare to the visual perception of human eye and saturation in the image sensor, hence makes it challenging for camera-based localization. The surface of a satellite can be partly diffuse and specular. A combination of the diffuse reflection of rough surfaces with the specular reflection of shiny surfaces is efficiently represented with the Phong reflectance model. The Phong reflectance model is a simplified BRDF model commonly used in computer graphics for rendering, particularly specular surfaces.

3.2 Camera and Lens System

A camera-based pose estimation and tracking along various ranges relies on a carefully selected camera and lens system. The resolution of the image, the sensitivity of the image sensor and characteristics of the lens of the camera, among many others, influence the performance of a 3D localization. In order to select and design an appropriate camera system for visual-localization, it is highly desirable to understand the basic principles of the image sensor and image formation process.

3.2.1 Camera Image Sensor and Optical Lens

An image sensor is a light sensitive electronic device that converts the light received on the camera into an electronic signal and records the image of a scene or an object. The

charge-coupled device (CCD) and complementary metal-oxide semiconductor (CMOS) are widely used image sensors. Both CCD and CMOS imagers convert light into electric charge and process it into electronic signals (voltage). A digital camera sensor is generally made up of three different layers; sensor substrate, a Bayer filter and a microlens. The sensor substrate is a silicon material, which measures the light intensity. The sensor has tiny cavities like wells, that collect the incoming light and allow it to be measured. Each well or cavity that appears flat is a pixel. The sensor is composed of millions of light sensitive areas or pixels. These light sensitive areas or wells can be thought of as a set of buckets, into which the light falls and is collected. The brightness level at each pixel is determined by the amount of light rays falling into each bucket. Once the bucket is full, the light level is saturated, resulting in maximum intensity (white saturated image).

The buckets measure only the intensity of the light. Specific light of a certain color can be captured by placing a different color primary filter on each well. The filter basically allows passing of specific light wavelengths. One specific color filter array is called Bayer filter. The Bayer filter acts as a screen, only allowing light photons of a certain color into each pixel on the sensor; each row or column of an image has only two of the three primary colors, either red and green or blue and green (Bayer pattern). In practice, pixels of an image are not precisely placed next to each other, therefore there exists a small gap between two pixels. Consequently any light that falls into this gap is wasted light. A microlens is used to enhance the ability of light gathering by directing the light that falls between two pixels into one or other of them. This increases the response (quantum efficiency) of the sensor particularly at shorter wavelength. This small lens is placed above the Bayer filter and helps each pixel to capture as much light as possible.

The Microlens, discussed earlier is an optional tiny lens placed on a gap between pixels to improve light capturing efficiency of the image sensor, however it does not replace the main optical lens of a camera. Most cameras use a lens to focus light onto the view plane to capture enough light in a sufficiently short period of time. This is a desirable property particularly since the objects may move appreciably, and the image needs to be bright enough to show significant detail over a wide range of intensities and contrasts. An optical lens is primarily characterized by its focal length, aperture and resolution which determine a field of view and magnification, amount of light entering to the image sensor and resolution of the image respectively.

Digital Images

The output of an image sensor, such as CCD is a continuous electronic signal, generated by scanning photo-sensors in a given order and reading out their voltages. The signal is sent to a frame grabber, which digitizes it into a 2D rectangular $N \times M$ array integer values, stored in the frame buffer. In a CCD camera, the physical image plane is the CCD array of $N \times M$ rectangular grid of photo-sensors.

The digitization process involves converting analog image data into a numerical representation through discrete sampling and quantization methods. Sampling converts continuous image signal into discrete data. The image signal is sampled at regular intervals as a grid of pixels to represent a digital image. The quantization process then assigns each sample a numerical value from a defined range (e.g. for 8-bit gray scale image, the range is 0 to 255). In image processing, this quantized values are often represented as an image intensity $I(x, y)$ at each pixel position (x, y).

3.2.2 Image Formation

Image formation depends on geometry and physics of light. The light determines brightness of a pixel in an image plane as a function of illumination and surface properties (Section 3.1). The geometry of image formation is used to determine where the projection of a 3D point is located in the image plane. Simply, when an object or a scene is illuminated by a single source, the scene or the object reflects radiation towards the camera and the camera senses it via photosensitive sensors (CCD or CMOS), as shown in Fig. 3.3. A pinhole camera model is used to describe the geometry of image formation.

Pinhole Camera Model

The pinhole camera model defines the geometric relationship between a 3D point and its corresponding 2D projection onto the image plane. It is the simplest model to form an image of a 3D object or scene on a 2D plane. The rays of light pass through a pinhole and form an inverted image of the object on the image plane. The pinhole camera approximates well the physical model of a real camera, enabling a linear mapping (in homogeneous coordinate) of a 3D to 2D between a scene and an image. Therefore,

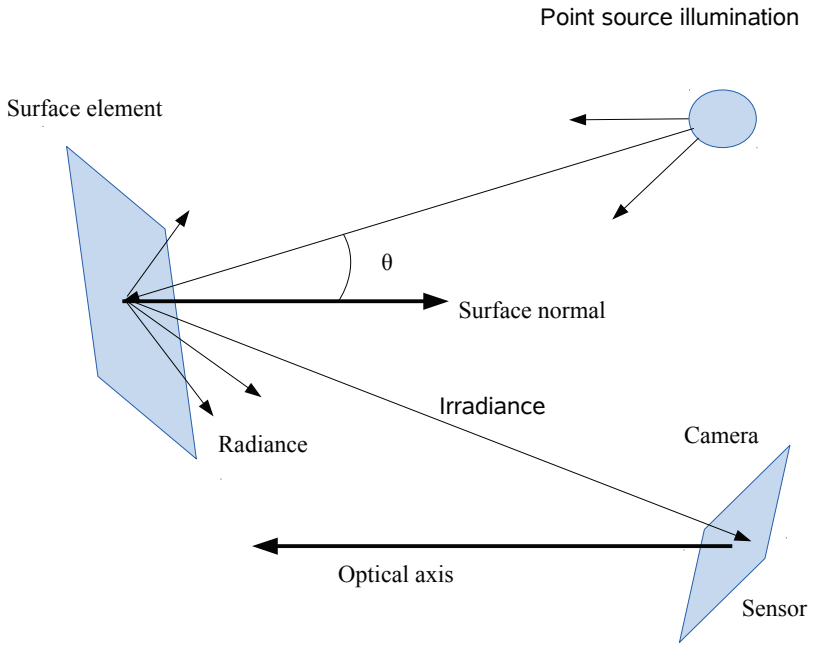


FIGURE 3.3: Image formation principle.

the pinhole camera model which is a perspective projection, is widely used in computer vision. The aperture is in practice larger a tiny hole (pinhole) to admit more light. Lenses are placed in the aperture to focus rays from each scene point onto the corresponding point on the image plane.

Given the scene points (X, Y, Z) and the image points (x, y), the perspective projection of a pinhole camera of focal length f is given by $x = f\frac{X}{Z}$ and $y = f\frac{Y}{Z}$. According to this perspective projection, the central projection of the pinhole camera is the linear mapping of the scene and image points in homogeneous coordinates, given by

$$\begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{3.1}$$

where the image points are $x = \frac{u'}{w'}$ and $y = \frac{v'}{w'}$. The origin of the image plane coordinate $(0, 0)$ is at the image center.

In practice, the origin of the pixel coordinate in an image is at the top left corner (Fig. 3.5). Therefore, the perspective projection of a 3D scene point onto image plane in pixel coordinates (x_{pix}, y_{pix}) is

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} f_x & s & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{3.2}$$

where the image points in pixel coordinate are $x_{pix} = \frac{u}{w}$ and $y_{pix} = \frac{v}{w}$, (c_x, c_y) is the principal point (image center), (f_x, f_y) is focal length in pixel and s is a skew parameter. Notice that the metric focal length f is converted to pixel using the scaling factors k_x and k_y, i.e. $f_x = fk_x$ and $f_y = fk_y$, where $k_x \times k_y$ is the resolution of an image sensor. The principal point and focal length can be estimated through a camera calibration.

3.3 Camera Calibration

A Camera calibration is a process of estimating the parameters of a pinhole camera model, which approximates the camera that produced a given image. The camera intrinsic parameters consists of the image center (principal point), focal length, skew factor and lens distortion parameters. The intrinsic camera parameters can be also represented in a matrix form as

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3.3}$$

where the upper triangular, the five degrees of freedom matrix K, is known as calibration matrix.

The camera calibration procedure reduces to the estimation of a 3×4 general matrix using scene points and corresponding image points, and estimate interior and exterior parameters from this matrix. This estimation can be achieved with a known geometry of a calibration plate (target object), which consists of several corners of a known position in the target coordinate frame, and corresponding images in different views.

If $\mathbf{p}$ is a surface point expressed in the object frame, then the general perspective projection in homogeneous coordinate onto an image plane is $\mathbf{u}$, such that

$$\mathbf{u} = K\Pi {}^{c}T_{o}^{o}\mathbf{p} \tag{3.4}$$

where K is a calibration matrix, ${}^{c}T_{o}$ is a homogeneous transformation matrix related to a rigid transformation from an object frame to a camera frame, and

$$\Pi = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3.5}$$

The camera calibration involves estimating the intrinsic parameters K and extrinsic parameters ${}^{c}T_{o}$. The calibration consists of solving a set of equations derived from correspondences of 3D scene points and image points for a general matrix $P = K\Pi {}^{c}T_{o}$. The matrix P is then decomposed to intrinsic camera calibration matrix K and extrinsic rotation and translation parameters. One of the most used calibration method in computer vision is one proposed by Tsai [145].

The pinhole camera model assumes a perfect lens system that enables to image lines to lines. In practice, the lens distorts the image particularly for cheap camera optics and short focal length cameras (wide field of view). A radial distortion is the main cause of image deformation, and can be modeled with a polynomial function [146],

$$\begin{aligned} x &= x'(1 + k_1 r^2 + k_2 r^4) \\ y &= y'(1 + k_1 r^2 + k_2 r^4) \end{aligned} \tag{3.6}$$

where (x', y') is the distorted point, k_1 and k_2 are distortion coefficients, and $r^2 = x'^2 + y'^2$.

More than one camera can be simultaneously calibrated. Calibration of two cameras in stereo configuration is common in computer vision. The stereo calibration is performed by taking images from two views at the same time. The intrinsic parameters of each camera and relative position and orientation of the second camera with respect to the

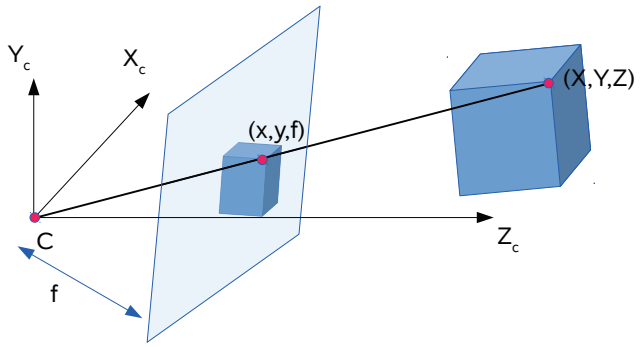


FIGURE 3.4: The scene point projected onto the image plane of pinhole camera, which is at a distance f from the center of projection C.

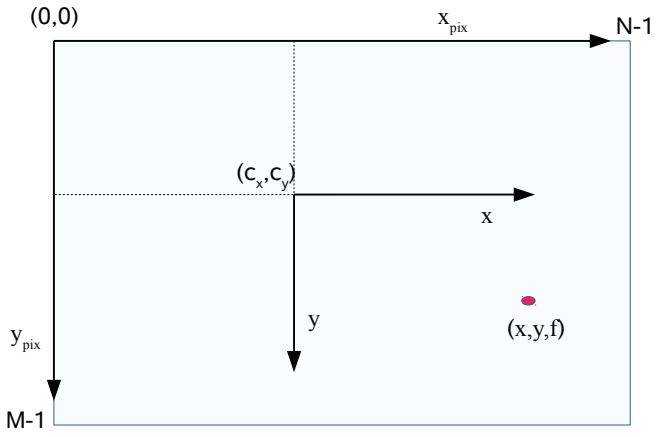


FIGURE 3.5: Image plane and pixel coordinate frames.

first is estimated based on synchronized images and a known geometry of the calibration plate. The camera calibration plate often consists of a checkerboard pattern and reference points (Fig. 3.6), which are precisely measured.

Hand-Eye Calibration

A camera calibration is necessary, for example to recover the geometry of a scene or tracking an object in 3D space. A hand-eye calibration is required whenever a camera is

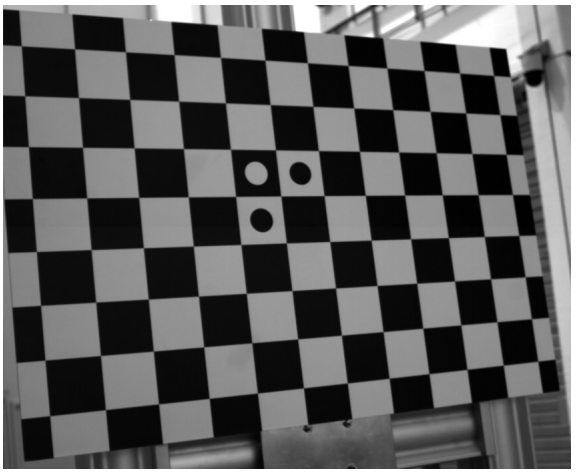


FIGURE 3.6: A calibration plate. The corner points of the checkerboard are extracted using a corner detector. The corresponding 3D points of the calibration plate are measured. The circles in the middle are used as a reference.

mounted on a robot hand (end effector). Determining the transformation between the camera and the hand-tool center point (TCP) of a robot is called hand-eye calibration. The hand-eye calibration is also necessary when the motion of a satellite is simulated with the robot motion (for detail, Chapter 4), consequently the associated ground truth can be used for the evaluation of vision-based motion estimation methods. Therefore, the relative orientation and position of the camera with respect to the TCP of the robot should be estimated through a hand-eye calibration procedure such as [147]. A common approach to estimate the camera sensor to robot TCP transformation is solving a matrix equation of the form $A_iX = XB_i$, where X is a 4×4 desired homogeneous transformation matrix (hand-eye), A_i and B_i, for i=1,..N are movements of the robot and induced camera motion respectively.

3.4 Local Appearance Cues

Local features of an image are robust to occlusion and clutter, and are suitable for a real-time tracking. The sparse distinctive set of features, which are to some extent invariant to a geometric and a photometric change are widely used for camera calibration, pose estimation and tracking. For this purpose, the image of an object is commonly represented by its features such as keypoints, edges and lines.

3.4.1 Image Features

Features are distinctive points or regions of an image despite change of view and lighting. The core idea in detecting keypoints and edges is to calculate gradients of an image at each pixel. There exist three possibilities for image regions around the pixel [36]. The image regions around the pixel can be:

- a flat region: both horizontal and vertical gradients are small, indicating no distinctive image feature at this pixel,
- an edge: one of the gradients (perpendicular to the edge) are big and the other (along the edge) is small; notice that line segments are also edges,
- a keypoint or corner: both gradients are big.

Keypoints

Keypoints are the most suitable image features for tracking. They exist on textured objects or at the intersection of segments in an image. There are several types of keypoint detectors. Some feature detectors are more suitable for tracking in video sequence because of their localization accuracy such as Harris [36] and Good features to track [6], others for real-time efficiency such as FAST [148]. There exist also a number of keypoint detectors more suitable for matching images with variable scale, for example SIFT [7].

Edges

Edges of an image are relatively robust to variation of illumination. The well known Canny edge detector [149] is often used for detection of edges in a given image because of its properties to reject noise and localize edges accurately (1-pixel precision). However, for object tracking edges such as straight line segments at a single orientation suffer from aperture problem, i.e. it is only possible to align the patches along the direction normal to the edge. Therefore, it is difficult to track an object based on edge features in the absence of the 3D model of the object. In contrast, edges are powerful visual cues in model-based tracking. The presence of the model of an object enables to efficiently search matches along 1D (along edge normal), as applied in Chapter 6.

3.4.2 Feature Descriptors

Scale and rotation invariance are desirable in several applications such as image matching and object recognition. Therefore, instead of finding stable and distinctive locations, it is more effective to extract features that are more stable in location and scale [7]. Lowe [7] proposed to look for a space-scale maximum in the structure computed from a set of sub-octave difference of Gaussian (DoG) filters. The scale selection criteria was based on the work of [150], who first used extrema in the Laplacian of Gaussian (LoG) function. Image matching and object recognition algorithms also need to deal with large image rotation.

The rotation invariance can be achieved by estimating a dominant orientation at each detected keypoint. The dominant orientation can be obtained by computing a histogram of orientations around the keypoint. This calls for necessity of feature descriptors around each keypoint. Feature descriptors are a set of feature vectors computed for an image at scaled and oriented patches around keypoints. Among feature descriptors SIFT [7] is widely used for its reliable image matching and object recognition. Fast feature descriptors include SURF [39] and also binary descriptors BRISK [42], BRIEF [41] and ORB [151].

3.5 Pose Estimation

One of the goals of this thesis is to estimate the rigid transformation between the client satellite and a camera based on the features extracted from camera images. The pose estimation in this context determines a coordinate transformation between an object and the camera coordinate frames. Closely related to pose estimation is motion estimation, which is the object motion between two time instants. In this thesis, motion estimation refers to pose estimation in the absence of a geometric model of the object.

3.5.1 3D-2D Alignment

Pose estimation from the correspondences between 3D points of an object and 2D points on image is based on a projection model. The common projection models used in pose estimation include full perspective projection, homography mapping, orthographic projection and weak perspective projection. With the full perspective model in Equation (3.1), 3D feature points are projected onto an image plane with perspective rays originating at the center of projection C, which lies within the physical camera as illustrated in Fig. 3.4. For planar surfaces, the non-linear constraint of the perspective projection reduces to a linear homography mapping. In orthographic projection, object points are perpendicularly projected onto the image plane neglecting any perspective distortion and scale of the object. Similar to the orthographic projection is a weak perspective projection, which estimates scale or overall distance. The orthographic and weak perspective models are valid when the depth of the object is very small with respect to the distance to the camera. In the following, we assume a more accurate full perspective projective model to describe the 3D-2D registration in image plane.

The pose estimation from a given set of 3D and 2D points is generally called Perspective-n-Point (PnP) problem. The PnP problem attempts to estimate [152–154] the orientation and position of a camera or an object, provided intrinsic parameters and a set of n correspondences between 3D points and their 2D projection. In order to reject spurious features in the correspondences, the robust RANSAC algorithm can be used with minimal sets of points for pose estimation (P3P problem) but at a cost of higher computational burden. The recently developed method [154] efficiently rejects outliers algebraically by back-projecting all points using estimated pose.

The pose estimation with unknown correspondences between 3D and 2D point sets is one of the most difficult problems in computer vision. This is a sub-problem of a global pose detection, which is addressed in Chapter 7. In the case of tracking, the problem of pose estimation with unknown correspondences may be relaxed, because the last estimated pose is used as a priori information for the current pose.

3.5.2 3D-3D Alignment

A 3D-3D registration usually arises when 3D data are obtained from range images such as time of flight cameras or stereo algorithms. The methodology to address the 3D-3D registration problem (estimation of a rigid transformation) depends on the relationship between the point sets. Iterative closest point (ICP) algorithm and its variant [155, 156] are commonly used to register the two 3D point sets which are irregular samples of the object with varying size. On the other hand, for the simpler case where the points sets are of equal size and known correspondences, the motion estimation can be achieved with [9, 45].

3.6 Geometric Models for Localization

A geometric model of an object such as a CAD model is useful for a vision-based localization, particularly the aperture problem associated with edge matching is efficiently exploited in the presence of the CAD model for the tracking of the object. Edges or line models of the object are used to estimate the camera pose by aligning the edges or line segments with edges found in the image.

3.6.1 Simple and Complex Satellite Models

The geometric model of a satellite may exist in two forms; simple 2D drawings of physical components and a CAD model. The 2D drawing describes the physical dimension of each component of the object. In this case, the 2D model drawing can be used either directly for tracking or is converted to a simplified 3D model. In addition to a CAD model, we use in this thesis a simple 2D drawing model of the engine nozzle and the boundary of a satellite for localization of an attitude-controlled satellite (Chapter 6). On the other hand, a CAD model is a 3D mesh used to design the object using computer graphics. The model may be wire-frame, surface or solid model. It may also consist of a texture model. The CAD model of a satellite (e.g. Fig. 3.7) is mostly complex, and cannot be used directly for tracking. In order to use the CAD model for a real-time tracking, some of the polygons (triangular meshes) have to be reduced by removing dense triangular meshes, while maintaining a few of them up to the desired accuracy. The wire-frame

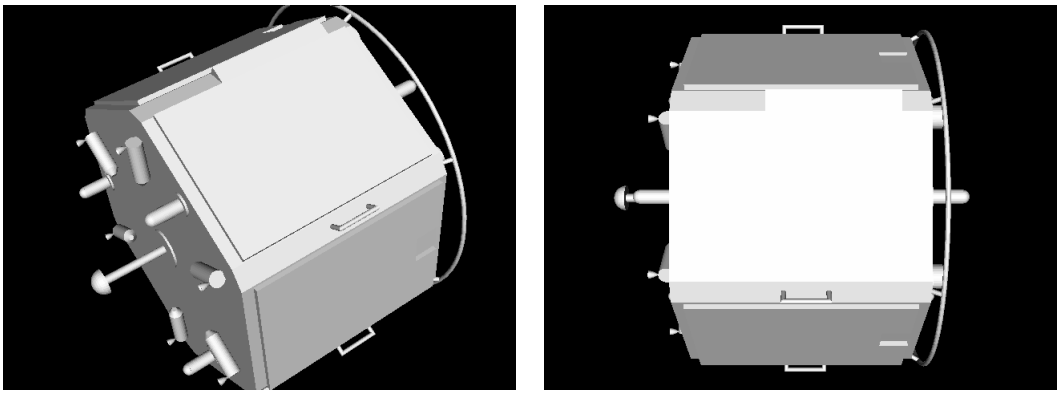


FIGURE 3.7: The 3D model of a satellite in 2D view from two view points.

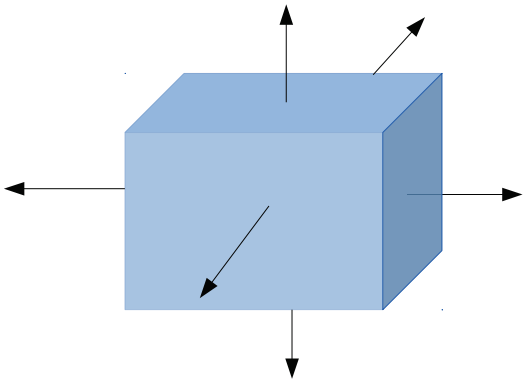


FIGURE 3.8: The normals of a model for edge-based tracking should point to uniform direction. The arrows indicate the direction of the normals, here all pointing outward.

model is used for the tracking, therefore most of the contours of the satellite are retained during the model pre-processing. Moreover, all the normals of the model should point to a uniform direction, either outward or inward for edge-based tracking as shown in Fig. 3.8. Notice that normals of polygonal faces and 3D features of the model such as edges are desirable for model-based tracking.

3.6.2 Model Features for Tracking

Image edges are relatively robust to illumination change, hence often used for model-based tracking. The model edges are used to match with the image edges to find the correct pose during tracking. Therefore, the model pre-processing pays attention to these edges and keep the most prominent contours. The model lines may be explicitly

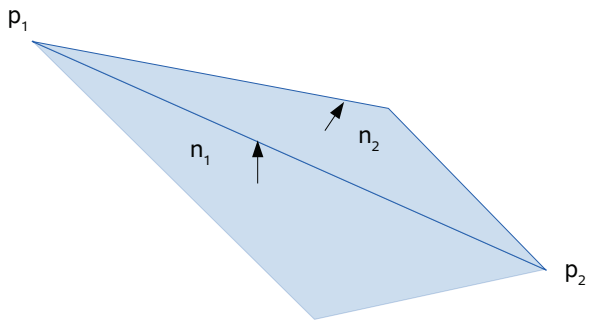


FIGURE 3.9: Model edge extraction. The model edge ($\overline{p_1 p_2}$) can be computed from the normals n_1 and n_2 of the polygons. An edge is defined when the angle between the two normals exceeds a certain threshold.

extracted from the CAD model by applying the Laplacian filter to the depth values computed with Z-buffer as in [157]. On the other hand, the angle between normals of two polygonal faces can be also used to determine the edge between the faces (Fig. 3.9). In this case, an edge is defined when the angle between the normals of the two faces is greater than a certain threshold.

3.7 Summary

Vision-based localization of an object depends on lighting, surface reflectance, image sensor, camera lens as well as the configuration and motion of an object. In order to localize the object in 3D space, a pinhole perspective camera model, a calibrated camera and the correspondences of image features and 3D features are widely used. Image features play a crucial role for the performance of localization as well as for camera calibration. Image features such as keypoints (e.g corners) and edges are used for image to image (2D-2D) or image to object (3D-2D) matching due to their robustness to occlusion, clutter as well as real-time efficiency.

The Perspective-n-Point (PnP) methods attempt to align the re-projected 3D features of an object to the corresponding 2D image features. This alignment results in an estimate of position and rotation (pose) parameters associated to the object motion with respect to the camera coordinate frame. On the other hand, a model-based tracking exploits model edges to align with image edges along the edge normal. Hence, model-based

edge tracking simultaneously searches for the 2D-3D correspondences and the pose by minimizing the re-projection error using local optimization methods such as Gauss-Newton. For real-time efficiency, the model of the object is pre-processed to reduce some details. The model edges are useful 3D features, which can be extracted either explicitly by filtering the Z-buffer or from the normals of adjacent faces.

Chapter 4

Ground Test Setup for On-orbit Servicing

Designing and implementing on-orbit servicing robotic space mission is a very complex and expensive endeavor. Hence it is absolutely necessary to perform on-ground validation and verification of subsystems before deployment. For the validation of camera-based tracking methods, we employ a ground based test facility, which simulates in-orbit space environment, satellite motion and its optical characteristics. The facility, which mainly serves for rendezvous and docking simulates:

- space environment such as illumination of the sun,
- optical characteristics of the satellite surface,
- motion of the satellite.

In particular, a visual tracking is validated using ground truth data, generated through hardware simulation systems and software (e.g rendering). Both methods may be able, albeit to a different extent, to provide realistic camera optics, motion of the client, sun illumination and optical characteristics of the client surface. On one hand, the software simulation relies on physical models, and is suitable for generating a trajectory with accurately known ground truth, despite the difficulty of simulating actual surface properties. On the other hand, hardware simulation (e.g. robotic) is more practical,

despite some ground-truth inaccuracies due to random and systematic errors of robot-based measurements and camera calibration, and less realistic sunlight and background illumination conditions.

Effect of Scaling the Client Mock-up

Scaling down the satellite size to fit to the maximum carrying capacity and workspace of the robot is often necessary. In this case, several factors should be incorporated in the analysis of the method. In particular, the field of view of the camera, which is designed for full-scale satellite tracking, introduces full satellite shape at close range. This in turn, enables the important satellite features which are not visible in full scale, will be available for good pose estimation. Thus, the reported result is far from reality. Furthermore, the dynamics are also affected as relative motion between servicer and target are scaled by some factor. In this study, we consider mock-ups of full scale satellite surface models for the validation of our algorithm, using robotic hardware facility and photo-realistic rendering. Moreover, our tracking algorithm does not rely on dynamic parameters such as inertia matrix and center of mass.

4.1 Hardware

We use the European Proximity Operation Simulator (EPOS), a robotic rendezvous and docking platform at DLR for ground-truth generation and testing [1, 158]. The EPOS facility reproduces a complete satellite rendezvous system, space environment and optical characteristics by using:

- mock-ups of servicer and client satellites,
- motion trajectory using two 6 DOF robots,
- multilayer insulation (MLI) on the satellite surface,
- high power floodlight with the sunlight spectrum (Fig.4.3).

FIGURE 4.1: Satellite mock-ups reproducing full scale satellite used in various chapters of this thesis: DLR satellite mock-up with golden foil (left) in Chapter 5 and the Kayser-Threde mock-up with real MLI (right) in Chapter 6 at the EPOS facility.

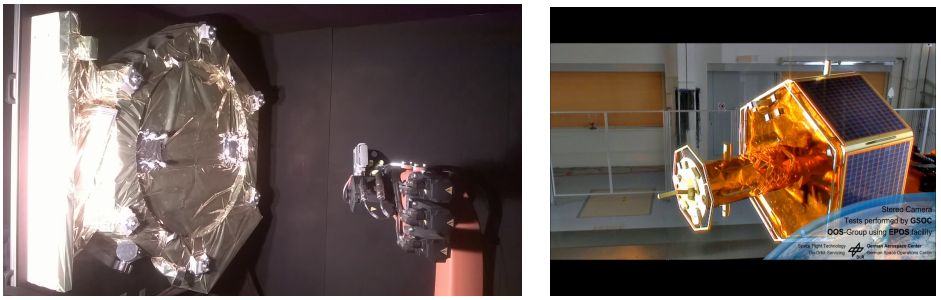


FIGURE 4.2: Satellite mock-ups reproducing full scale satellite used in Chapter 7: Rear part of TerraSAR-X satellite with reflective foil (left) and Airbus DS GmbH satellite mock-up (right).

4.1.1 Satellite Mock-ups

We consider various mock-ups, serving for different purposes. Firstly, we focus on the rendezvous of the client satellite which is predominately specular surface due to reflective golden foil (Fig.4.1) which reproduces MLI. This satellite mock-up (Fig.4.1 left) is used to evaluate tracking methods, for the spinning motion of a satellite around its major axis under strong specular reflection. Furthermore, a similar satellite mock-up but with larger size and real MLI provided by the company Kayser-Threde GmbH (Fig.4.1) is used to track position of an attitude-controlled satellite, handling small attitude fluctuations. The latter, is used to evaluate a satellite nozzle detection and tracking method. In this case, external guidance, navigation and control sensors such as star trackers are assumed available, and used to align servicer and client attitudes at predetermined range.

Moreover, in order to validate and verify vision based methods at very close-range, the rear part of a full scale TerraSAR-X Earth observation satellite (Fig.4.2 left) is built at the Robotics and Mechatronics Center, DLR. This mock-up consists of various structural

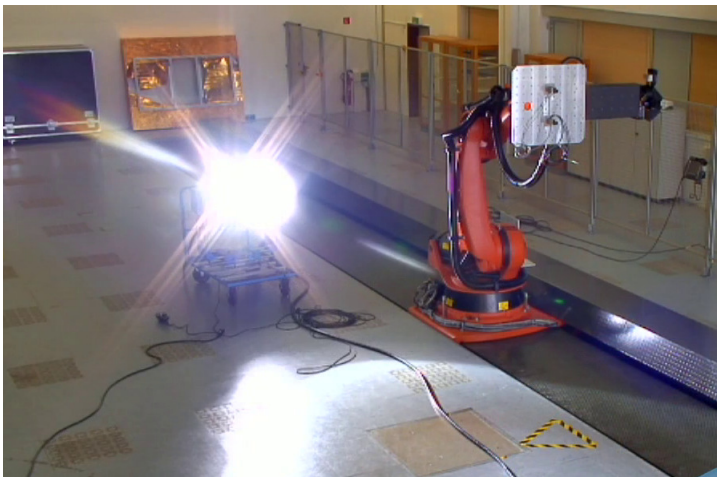


FIGURE 4.3: High-power floodlight simulates the sun at the EPOS facility.

elements and MLI, and can be used to assess camera based pose estimation and tracking methods. On the other hand, the full tumbling motion of a satellite is considered in this thesis. For this purpose, a DEOS client mock-up built by Airbus DS GmbH (Fig.4.2 right) is at the EPOS facility.

4.1.2 Optical Environment

The sun spectrum cannot be accurately reproduced with hardware. Furthermore, the source of light cannot be placed at infinity in practice. However, the characteristics of the illumination condition can be reproduced. In practice, high power floodlights provide sufficient spectrum to simulate space environment. In fact there exist other sources of light from distant stars and planets, but they affect the optical environment of space insignificantly.

4.2 Photorealistic Rendering

In addition to hardware-based simulation of space environment and satellite for visual tracking, we employ a rendering software (developed by SpaceTech GmbH) that produces realistic images through ray tracing. It consists of stereo cameras with a baseline of 40cm, a simple servicer satellite model, and a complex, photo-realistic client model. The program simulates camera optics, material properties of the client and sunlight spectrum. Both satellites are at 400 km altitude, and the Earth is modeled as a sphere

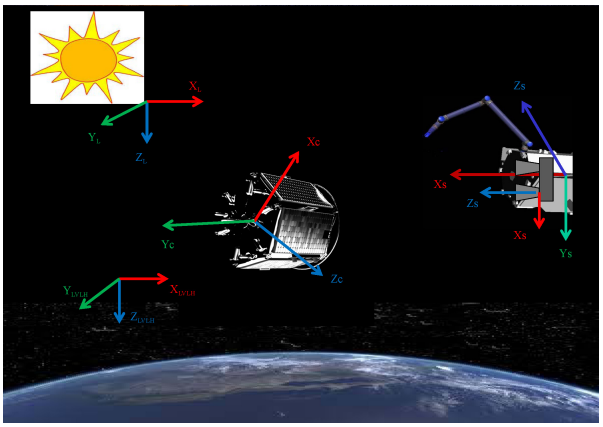


FIGURE 4.4: Photo-realistic rendering setup simulating servicer, client and the sun used in Chapter 7.

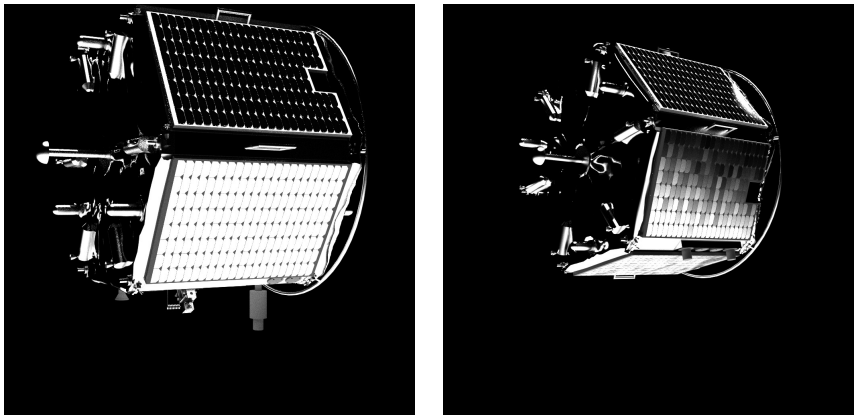


FIGURE 4.5: Renderered images

with constant albedo of 0.3 (Fig. 4.4). Since ray tracing is computationally intensive, we employed cloud computing parallelism using a computing cluster, to implement a desired trajectory in a relatively short time. The outputs of the renderer from each cluster computer (partial solution of the rendering equations) are collected from each cluster computer and averaged, to produce the desired image quality (Fig. 4.5). Each renderer produces 64 samples/pixel. The render program accepts as inputs the absolute poses of the servicer and client, the sunlight direction, the desired number of bins (that determines the image quality) and the integration time of the sensor (shutter time). In order to render a complete trajectory, we generate a script that consists of a given number of bins, and a set of poses. This script is executed in a batch; as a result, each camera view is rendered in separate nodes of the cluster.

4.2.1 Optical Environment and Satellite MLI Simulation

The simulation of the space environment in computer graphics, unlike that of the hardware-in-the-loop system, is relatively easy. The main components that build the environment is creating the sun's illumination and shadows. However, it is hardly possible to reproduce MLI of the satellite in computer graphics accurately. In general there exist two methods to create MLI model; 3D scanning of the real MLI and mapping to the geometric model, and imaging the real MLI with camera and texture mapping to the model. The disadvantage of these two methods are that scanning and imaging the MLI, which is highly reflective, do not capture the surface property properly. Nevertheless it represents approximately the reflective characteristics of the MLI.

4.2.2 Satellite Motion

One of the advantages of computer graphics simulation is that the motion of the satellite can be fully realized without any constraint. In contrast, motion of the client satellite in hardware simulation systems such as robots is restricted with robot workspace and joint limits. We have the capability to simulate the actual satellite motion, such as spinning and tumbling about any of its axes, freely in time and space.

4.3 Summary

Realistic space environment and optical characteristics of a satellite needs to be simulated on-ground for validation and verification of a vision-based tracking and motion estimation. The simulation system should reproduce at least three basic elements: the sunlight, motion of the satellite, optical characteristics of the MLI as well as satellite geometric structures. In this thesis, the space environment and optical characteristics of the satellite are reproduced with both, hardware and software system.

The hardware simulation consists of two six DOF robots to reproduce the relative motion of the servicer and client satellites, a high power floodlight to simulate the Sun and a full scale client satellite mock-up with multilayer insulation. Although a hardware-based simulation system provides an accurate representation of the real system, it has a few drawbacks. The first drawback is that the floodlight as a non-distant light source unlike

the Sun emits non-parallel light rays near the sun spectrum, and its intensity diminishes against distance quadratically. However, this drawback can be compensated to some extent by moving the light source according to the range of the experiment. The second drawback is, inaccuracy of the ground truth used for evaluation as the result of the robot positioning error (about 4 mm), which is generally acceptable for the task of on-orbit servicing.

On the other hand, a computer graphics based simulation reproduces the sunlight illumination and the ground truth trajectory more accurately. In spite of achieved photo-realistic camera images, reproducing the optical characteristics of MLI through rendering is difficult and less accurate compared to the hardware simulation. Therefore, it is an essential that both hardware and software based test bed should be employed in order to exploit strengths and avoid weaknesses of each method.

Chapter 5

Model-free Motion Estimation under Space Lighting

The geometric model of an object provides useful a priori information for accurate and robust motion estimation. However, the CAD model may not exist or is inappropriate due to various reasons. Although the satellite model is usually available, there are certain cases that it can be lost or badly handled. In particular, lack of geometric information of fragmented satellite debris and asteroids is quite imminent. Therefore, it is necessary to provide a 3D motion tracking method to address such scenario for on-orbit servicing. In this chapter, we propose and demonstrate a robust feature-based 3D tracking method in the absence of a geometric model. Although image edges are generally robust to change of illumination, they are difficult to track in the absence of the geometric model of an object because of the aperture problem.

5.1 Predictive Rigid-body Motion Tracking

Feature-based tracking methods have been successfully employed for various applications (for example, robot navigation, industrial object picking) under slightly variable lighting conditions. Under this condition of illumination, keypoints such as corners [6] can be selected in the first frame and tracked over the next frame (sparse optical flow tracking), with a multiresolution implementation satisfying most of the required real-time efficiency and accuracy. Since the above mentioned work, several extensions appeared. A few of

them incorporated photometric and geometric cues of a scene [159], improving computational efficiency through an inverse compositional approach [160, 161]. Among the optical flow based tracking methods, [144] incorporated specular highlights and brightness variations by reformulating brightness constancy assumption. On the other hand, instead of utilizing extracted features directly for tracking, a scale and rotation invariant feature descriptor such as SIFT [7] is used to describe each incoming image. Then, the tracking proceeds by matching descriptors of each frame based on geometric distance. The later approach is computationally demanding, therefore it may be currently inappropriate for real-time tracking.

On the top of such image-level tracking, 3D relative pose tracking is achieved by introducing more than one camera so that spatial feature correspondences across the second view can be established to retrieve 3D spatial information. Thus, the standard feature-based approach to 3D tracking is to triangulate tracked and matched points, and register corresponding 3D points using relative pose estimation algorithms, based on SVD [10, 44], quaternions [45] or dual quaternions [9].

This chapter is based on our work [162, 163], which is motivated by the existing correlation between image flow and disparity as in [164], and stereo-motion fusion, using the ratio of disparity rate to disparity, for recovering the rigid motion parameters [165]. We show a novel method for 3D sparse structure recovering and tracking, that exploits keypoints of a rigid object (satellite), with the capability of predicting features in absence of image measurements. Prediction capability is particularly essential since sufficient image features may not exist because of sensor saturation, shadow and unstable specular features of the satellite. The approach is based on a bank of iterated, extended Kalman filters (IEKF), using point-wise kinematic models, therefore not requiring any model of global target geometry or dynamics.

5.1.1 Point-wise Kinematics of Rigid-body Motion

The motion of 3D surface points $\mathbf{p} = [X, Y, Z]^T$ of a rigid body, translating and rotating with velocity $\mathbf{V}$ and $\boldsymbol{\omega}$ in camera frame respectively, is expressed by

$$\dot{\mathbf{p}} = \mathbf{V} + \boldsymbol{\omega} \times \mathbf{p} \tag{5.1}$$

while the projection onto camera pixel coordinates, through a pinhole model, is given by

$$\begin{aligned} x &= f_x \frac{X}{Z} + c_x \\ y &= f_y \frac{Y}{Z} + c_y \end{aligned} \tag{5.2}$$

where f_x, f_y are the focal length, and c_x, c_y the principal point coordinates in pixels.

By taking the total time derivative of Eq.(5.2) and substituting into Eq.(5.1), the motion field and rigid body velocities are related by

$$\mathbf{u} = \mathbf{M}^T \boldsymbol{\Omega} \tag{5.3}$$

where $\mathbf{u} = \begin{bmatrix} u & v \end{bmatrix}^T$ is the feature velocity in the image plane

$$\mathbf{M} = \begin{bmatrix} -f_x/Z & 0 \\ 0 & -f_y/Z \\ (x-c_x)/Z & (y-c_y)/Z \\ (x-c_x)(y-c_y)/f_y & f_y(1+((y-c_y)/f_y)^2) \\ -f_x(1+((x-c_x)/f_x)^2) & -(y-c_y)(x-c_x)/f_x \\ f_x(y-c_y)/f_y & f_y(x-c_x)/f_x \end{bmatrix}$$

and

$$\boldsymbol{\Omega} = \begin{bmatrix} V_x & V_y & V_z & \omega_x & \omega_y & \omega_z \end{bmatrix}^T \tag{5.4}$$

is the velocity vector.

At each frame detected features are tracked over the sequence; consequently, the rigid body velocity, which is used in the prediction stage of the Kalman filter, is estimated by minimizing

$$\underset{\boldsymbol{\Omega}}{\operatorname{argmin}} \left\| \mathbf{M}^T \boldsymbol{\Omega} - \mathbf{u} \right\|^2. \tag{5.5}$$

Estimating Depth Z

The matrix M in Equation (5.4) consists in an unknown parameter Z at each feature point, which can be determined in two possible methods. The first method is obviously matching across stereo views to estimate depth at each feature point. The prediction phase of extended Kalman filter depends on the accuracy of the estimated rigid body

velocity, which in turn relies on estimated depth. Therefore, so as to estimate depth as accurately as possible, every few frame new features are detected in one view and matched in the second view. This regular detection guarantees an accurate depth estimate in spite of insignificantly increased computation time. The alternative to this approach on the other hand, is to exploit tracked features over the sequence as well for stereo matching. However, the latter approach is inaccurate and not robust to specular reflection.

The proposed second method to depth estimation assumes a constant velocity model between two consecutive frames. Under such assumption, the last velocity from previous frame is used to propagate the 3D points to the current frame. Hence the current a priori state, predicted from the last state using the estimated rigid body velocity, is combined with current measurement (tracked features) to refine the posterior state estimate. At the current frame and filtered 3D points, the current velocity is estimated to be used for the state model in order to propagate the next 3D points. This method, unlike the first approach, can be used with a monocular camera although it requires initialization.

5.1.2 Iterated Extended Kalman Filter

The Kalman filter is a family of the Bayesian estimator commonly used for state estimation, filtering, and prediction. The general Bayesian estimator is

$$p(\mathbf{s}_k|\mathbf{z}_k) \propto p(\mathbf{z}_k|\mathbf{s}_k) \int_{\mathbf{s}_{k-1}} p(\mathbf{s}_k|\mathbf{s}_{k-1}) p(\mathbf{s}_{k-1}|\mathbf{z}_{k-1}) d\mathbf{s}_{k-1} \tag{5.6}$$

with state statistics $\mathbf{s}_k$, measurements $\mathbf{z}_k$, and the last posterior distribution $p(\mathbf{s}_k|\mathbf{z}_k)$ is determined through state dynamics and measurement likelihood.

The predictor-corrector filter arises from the Bayesian equation(5.6) is split into

- predictor:

$$p(\mathbf{s}_k|\mathbf{z}_{k-1}) \int_{\mathbf{s}_{k-1}} p(\mathbf{s}_k|\mathbf{s}_{k-1}) p(\mathbf{s}_{k-1}|\mathbf{z}_{k-1}) \tag{5.7}$$

- corrector:

$$p(\mathbf{s}_k|\mathbf{z}_k) \propto p(\mathbf{z}_k|\mathbf{s}_k) p(\mathbf{s}_k|\mathbf{z}_{k-1}) \tag{5.8}$$

The Kalman filter is a specialized form of the Bayesian estimator with certain assumptions on dynamics and likelihood. In particular, the Kalman filter addresses state estimation problem, by assuming linear process dynamics and normally distributed independent process and measurement noise. If all noise is Gaussian, the Kalman filter minimizes the mean square error of the estimated parameters, resulting in optimal state estimation.

The state $\mathbf{x}$ of a process that is governed by a stochastic difference equation of the form:

$$\mathbf{x}_k = A_k \mathbf{x}_{k-1} + B_k \mathbf{u}_k + \rho_k \tag{5.9}$$

with measurement $\mathbf{z}$

$$\mathbf{z}_k = H_k \mathbf{x}_k + \nu \tag{5.10}$$

where A_k is a matrix relating the previous state $\mathbf{x}_{k-1}$ with the current state $\mathbf{x}_k$, B_k is a control input of the system, and ρ is the process noise. H_k is related to the current state, with the measurement $\mathbf{z}_k$ while ν is measurement noise.

The Kalman filter provides an optimal state estimate. However, when the system model (either process or measurement model) deviates from linearity as it occurs in most real world problems, only a sub-optimal state estimate is guaranteed using the variant of Kalman filter (an extended Kalman filter). The extended Kalman filter linearizes model and measurement equations about an estimate of the current mean and covariance. In contrast, in our problem where the measurement equation introduces a non-linear system, iterated extended Kalman filter outperforms EKF because it uses the measurement to linearize the measurement function. Thus, after the prediction stage the iteration starts by first setting for i=0, $\mathbf{x}_k^0 = \mathbf{x}_{k,k-1}$ and $\mathbf{x}_k^{i+1} = \mathbf{x}_{k,k}$

$$H_k^i = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_k^i} \tag{5.11}$$

where $\mathbf{h}$ is a non-linear measurement equation relating to the state and the measurement. As a result, the Kalman gain, state covariance and state are updated respectively as

$$\begin{aligned} K_k^i &= P_{k,k-1} H_k^{i^T} (R + H_k^i P_{k,k-1} H_k^{i^T})^{-1} \\ P_{k,k-1} &= P_{k,k-1} - K_k^i H_k^i P_{k,k-1} \\ \mathbf{x}_k^{i+1} &= \mathbf{x}_k^i + K_k^i (\mathbf{z}_k - \mathbf{h}(\mathbf{x}_k^i)). \end{aligned} \tag{5.12}$$

5.1.3 Integrating Point-wise Kinematics and Image Motion

Assuming a small rotation of the rigid-body, a single-point kinematics (Eq.5.1) at discrete time k can be represented by

$$\mathbf{p}_k = \mathbf{A}_k \mathbf{p}_{k-1} + \mathbf{B}\mathbf{V}_{k-1} + \rho_k \tag{5.13}$$

with

$$\mathbf{A}_k = \begin{bmatrix} 1 & \Delta t\omega_{zk} & -\Delta t\omega_{yk} \\ -\Delta t\omega_{zk} & 1 & \Delta t\omega_{xk} \\ \Delta t\omega_{yk} & -\Delta t\omega_{xk} & 1 \end{bmatrix} \tag{5.14}$$

$$\mathbf{B} = \begin{bmatrix} -\Delta t & 0 & 0 \\ 0 & -\Delta t & 0 \\ 0 & 0 & -\Delta t \end{bmatrix} \tag{5.15}$$

$$\mathbf{V}_k = \begin{bmatrix} V_{xk} & V_{yk} & V_{zk} \end{bmatrix}^T. \tag{5.16}$$

The measurement model is

$$\mathbf{z}_k = h(\mathbf{p}_{k/k-1}) + \nu_k \tag{5.17}$$

where Δt is the time interval between two consecutive image captures and $\mathbf{z}_k$ are feature pairs of the stereo camera, tracked over the sequence and reprojected under the pin-hole model (5.2). Measurement and motion model errors, ν_k and ρ_k respectively are assumed Gaussian zero-mean additive vectors.

$$\begin{aligned} p(\rho_k) &\approx N(\mathbf{0}, Q) \\ p(\nu_k) &\approx N(\mathbf{0}, R) \end{aligned} \tag{5.18}$$

where Q and R are the process and measurement noise covariances respectively. For rectified cameras, with baseline T_x, we have

$$\mathbf{h} = \begin{bmatrix} f_{xl}X/Z + c_{xl} \\ f_{yl}Y/Z + c_{yl} \\ f_{xr}(X + T_x)/Z + c_{xr} \\ f_{yr}Y/Z + c_{yr} \end{bmatrix} \tag{5.19}$$

For the rectified stereo cameras the y coordinate is common, hence one element of $\mathbf{h}$ is redundant and can be discarded. The Jacobian H_k is computed as

$$H_k = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{p}}\right|_{\mathbf{p}=\mathbf{p}_{k|k-1}} \tag{5.20}$$

$$\frac{\partial \mathbf{h}}{\partial \mathbf{p}} = \begin{bmatrix} f_{xl}/Z & 0 & -f_{xl}X/Z^2 \\ 0 & f_{yl}/Z & -f_{yl}Y/Z^2 \\ f_{xr}/Z & 0 & -f_{xr}(X+T_x)/Z^2 \\ 0 & f_{yr} & -f_{yr}Y/Z^2 \end{bmatrix} \tag{5.21}$$

where, l and r indicate left and right stereo camera, respectively.

For each point, we iterate the updated stage of the filter until convergence. The main advantage of Bayesian filtering here is that, when a complete occlusion occurs, or the target moves momentarily out of the field of view, motion can be still predicted and used later on to re-initialize the tracking. Furthermore, at each step features may be selected independently of those chosen from the previous image, for dynamic occlusion handling.

The pose is finally estimated based on the dual quaternion [9], which is fast and numerically stable. Tukey M-estimator along with median statistics is applied to tracked 3D points to reject spurious features. This robust M-estimation allows to provide the tracking system, the available inliers for pose estimation and to signal the failure detection in case of insufficient inlier features.

5.2 Evaluation

In the following, the experimental results of the method described in Section 5.1 is presented and discussed. In this study, the extreme conditions such as strong reflection and shadow are stressed out. We consider the critical phase of a close-range approach (0.6 m to 1.5 m), where accuracies of less than 1° in rotation and 10 mm in translation are required. The goal of this experiment is thus to demonstrate and assess the feature based 3D tracking method for motion estimation of the satellite under harsh illumination condition and in the absence of a geometric model.

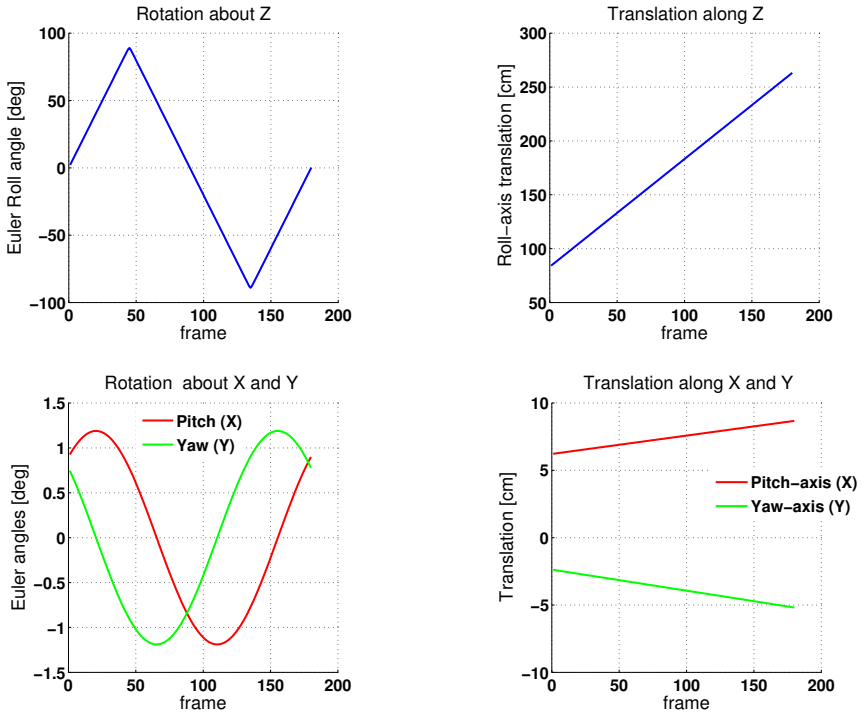
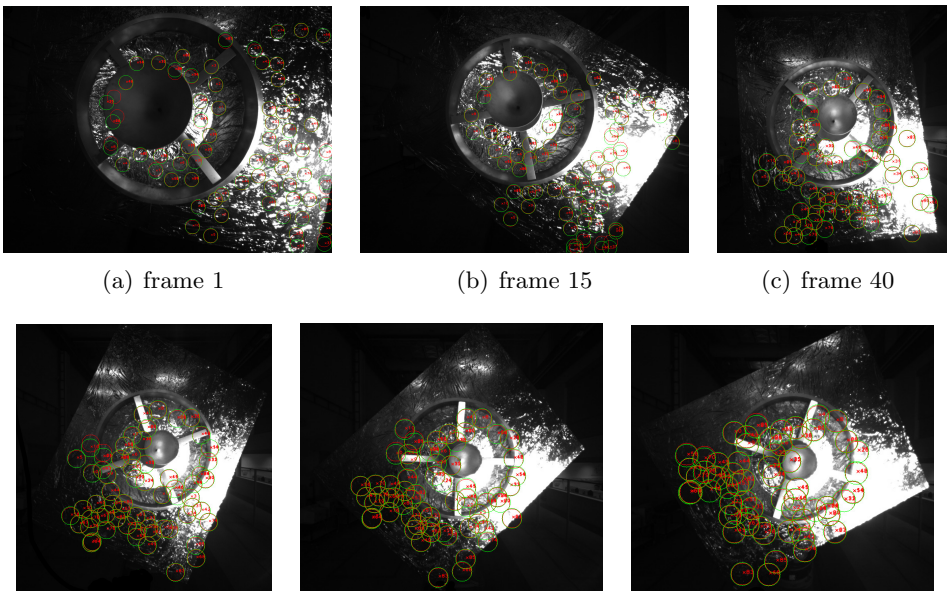


FIGURE 5.1: Typical ground truth trajectory for testing the motion estimation algorithm. The target spins at angular velocity of 2 deg/s and linear velocity of 10 mm/s

In the implementation point of view, a gray-scale image is acquired with a progressive scan CCD Camera sensor at a resolution of 1024×768 pixels and a field of view $87° \times 71°$. For the stereo based method, a baseline of 120 mm is employed. The main parameters that affect convergence and efficiency are the number of iterations in IEKF. The algorithm converges to the solution in 10 iterations. The measurement and process noise covariances are set constant over all the sequences. In order to evaluate and compare performances, a standard 3D stereo tracking method is also implemented with the same parameter setting of our method.

To study the effect of motion on the tracker, various translation and rotation speeds are considered; a typical trajectory is given in Fig. 5.1. Furthermore, the effect of nutation rate on the tracker is investigated. For purpose of evaluation of the method, we analyze frame-to-frame pose estimation and as well as the poses referred to the first pose of the target. The latter is used to analyze the drift of the features. Moreover, we compare to a standard stereo tracking method as well as to a monocular camera based method.



(a) frame 1 (b) frame 15 (c) frame 40

(d) frame 55 (e) frame 66 (f) frame 74

FIGURE 5.2: Some frames of the image sequences, with the client motion at a velocity of 2 deg/s and 10 mm/s and nutation angle of 5°. The overlapping circles (red and green) centered on the re-projected and tracked feature points respectively indicate successful matching.

5.2.1 Results and Discussion

The first set of experiment is to evaluate the performance of the method with respect to the ground truth poses under various motion trajectories. The image sequences as can be observed in Fig. 5.2 are very challenging for camera-based algorithms, depending on the reflection and shadow (dark) due to the multilayer insulation. Tracking features over long sequences is very limited, since detected features disappear even if the target is in frontal view. In this context, tracking each feature with its kinematics model in 3D helps predict the motion. We note here that the accuracy of the motion model depends on the estimated depth from stereo matching which provides object velocity. Thus inaccuracy in the motion model is mainly carrying from errors in depth estimation. In fact, the deviation of the kinematics model is modeled as Gaussian noise to represent uncertainty.

Observing the overall distribution of the error (Fig. 5.3) using frame-to-frame motion tracking, it indicates the feasibility of the method, at least in this critical range. This can be noted with very low errors in Fig. 5.3. Relatively large motion was around roll-axis (Z) in which the client was spinning at an angular and translational velocity of 2

deg/s and 1 cm/s respectively.

On the other hand, the translation and rotation error provided by a monocular camera are comparable to that of the stereo camera as can be observed in Fig. 5.4. The monocular camera can be employed for slightly less accuracy and robustness than that of stereo, however the monocular camera depends on good initialization. In fact, in space system hardware redundancy is usually employed, hence we can take advantage of such redundant camera view to initialize the monocular camera tracking.

On the other hand, standard stereo 3D tracking provides comparable accuracy as that of our method for some frames in the test sequence (Fig. 5.7 and at relatively higher velocity in Fig. 5.8), except in situations where the image saturation is very strong as observed in Z-translation in Fig. 5.7. While our method is as accurate as the standard stereo method, it is robust due to the prediction capability in case of dominant specularity and mild image saturation. In this case, the advantage of our method (new stereo method) can be noticed in Fig. 5.5, where rotation and translation errors of the main longitudinal axis are shown. Here, the new method is able to predict the motion for a few seconds (25 frames, up to a distance of 25 cm) when a few or no image features exist (Fig. 5.6).

The known drawback of model-free motion estimation is drift of features from the correct location because of accumulation of small errors. The drift is reduced by predicting feature location, using point-wise kinematics model (Fig. 5.9). To verify that as shown in this graph, features are detected in the first frame and tracked over the sequence, while each 3D point is registered with the corresponding first frame points. The drift is predominantly in spinning axis and camera optical axis (roll-axis) along the longitudinal motion. We remark that no feature rejection and replacement are performed during the 2D tracking, as we need to estimate the poses relative to its initial position, where the first frame has been acquired. However, spurious features are effectively rejected at the higher level, by the robust pose estimator. When the pose estimator runs out of the minimum required number of points, a new reference frame is taken, its keypoints are detected, and the resulting pose estimations are performed relative to the new frame.

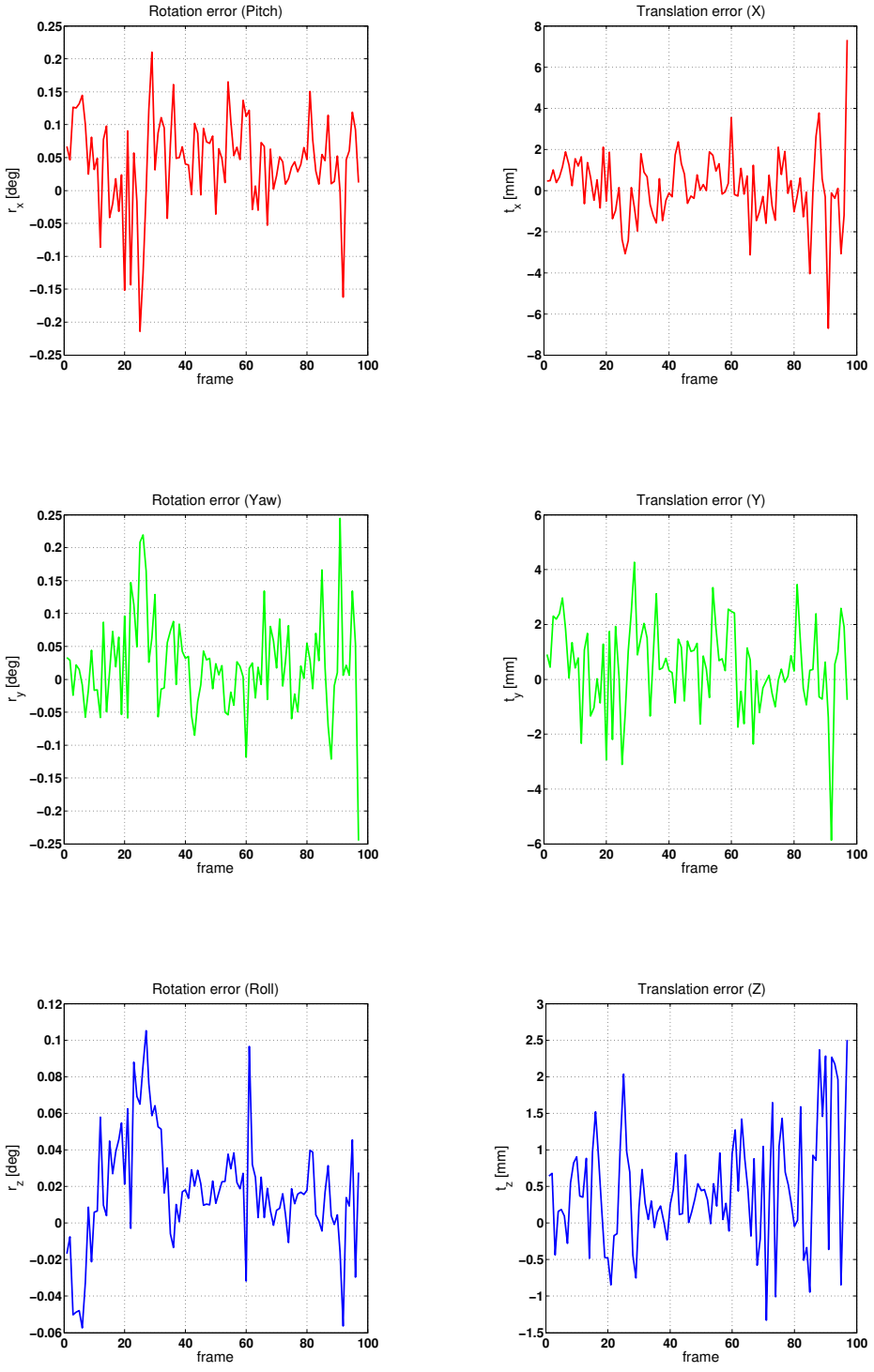


FIGURE 5.3: Rotation and translation error of the developed method as evaluated to ground truth. The target spins at angular velocity of 2 deg/s and linear velocity of 10 mm/s and frontal to the camera.

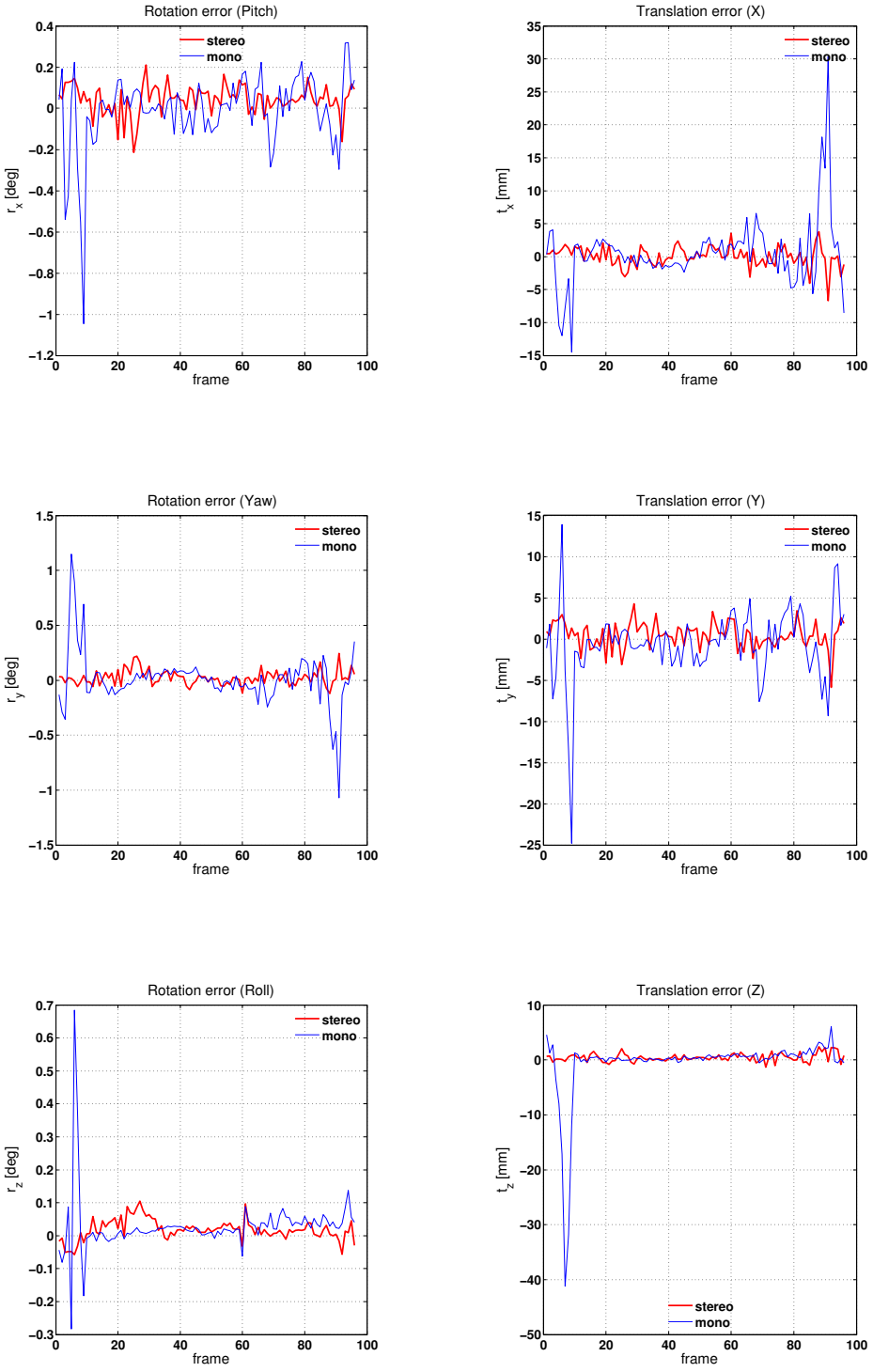


FIGURE 5.4: Comparison of stereo and monocular camera based methods, when the target spins at angular velocity of 4 deg/s and linear velocity of 10 mm/s and frontal to the camera.

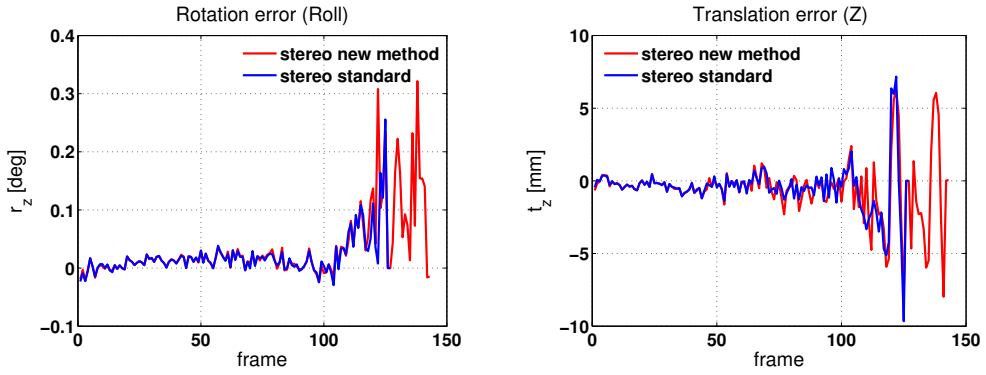


FIGURE 5.5: Predictive motion estimation based on a point-wise kinematic model. The estimated last velocity is used also in case of insufficient features to predict the motion. The new method is able to predict the motion accurately.

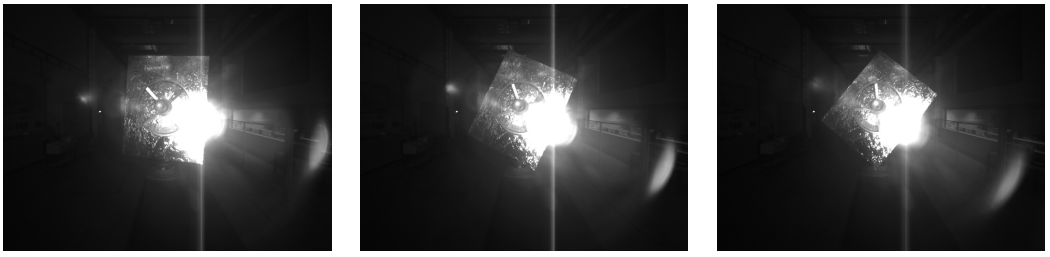


FIGURE 5.6: A few frames of the error plot of the predicted motion in Fig. 5.5. The features in these images are unstable for tracking because of the glare and saturation of the camera sensor.

5.2.2 Summary

In the absence of the CAD model of a satellite, only a relative frame-to-frame motion can be estimated. The accuracy of the relative motion estimation depends on distinctive image features. For the tracking of a satellite at close range, the wrinkle of the multilayer insulation and corners (intersection of satellite structures) provide useful cues for an accurate motion estimation and tracking. However, sufficient features may not always exist due to harsh space illumination (poor lighting and over-illumination), hence the tracking is prone to sudden failure.

To alleviate this problem, the thesis presented a predictive motion model that enables to track the motion even under worst illumination. For this purpose, the 3D points related to visual image features are predicted based on a point-wise kinematic motion model. The prediction of sparse 3D points is based on an accurate motion model which relies on the linear and angular velocity, which in turn depends on accurate estimation of depth at each feature point. The depth estimation is obtained from stereo cameras,

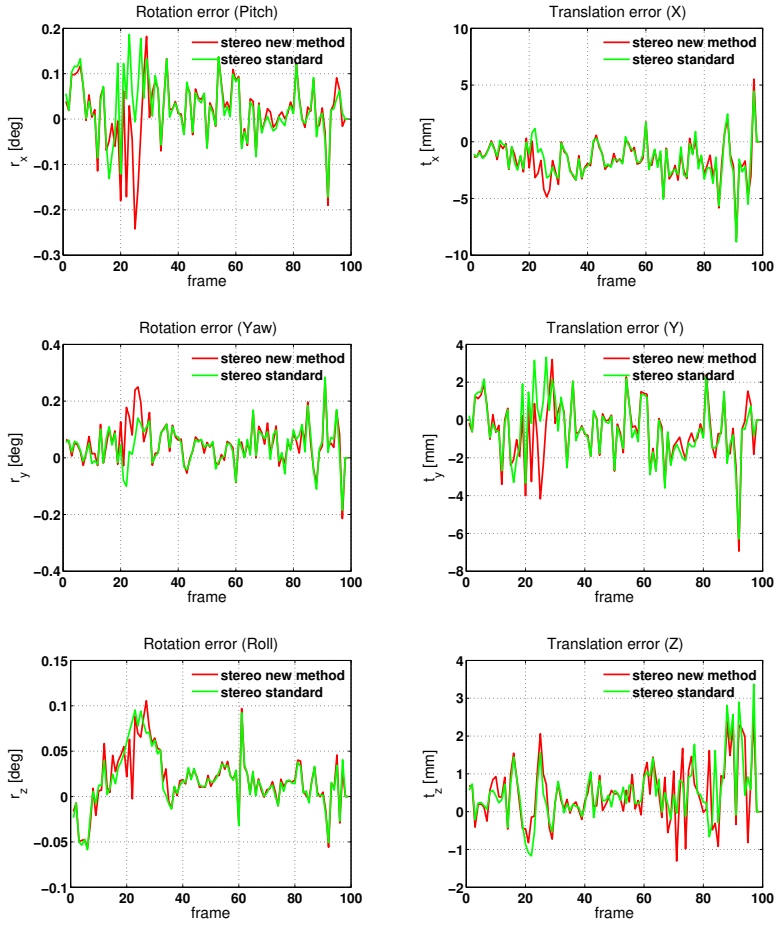


FIGURE 5.7: Comparison of the proposed stereo (new) and standard stereo tracking methods, when the target spins at angular velocity of 2 deg/s and linear velocity of 10 mm/s and, is frontal to the camera. The new stereo tracking method outperforms the standard stereo method when insufficient features exist, as can be observed on the graph of translation error along spinning axis (Z).

which is sufficiently accurate for close range approach and manipulation. In case of a failure of one of the camera electronics due to radiation, our method is able to proceed the tracking under the assumption of constant frame to frame velocity.

The motion tracking is demonstrated to be an accurate and more robust using a realistic satellite mock-up, motion trajectories and a high power floodlight to simulate the space lighting. The prediction capability of the tracker is, however, limited to a short-term scenario where under-and over-saturation of the image due to poor or strong lighting is tolerated for a few seconds. Moreover, an abrupt change of motion particularly with monocular tracking that relies on constant frame-to-frame velocity, is not guaranteed.

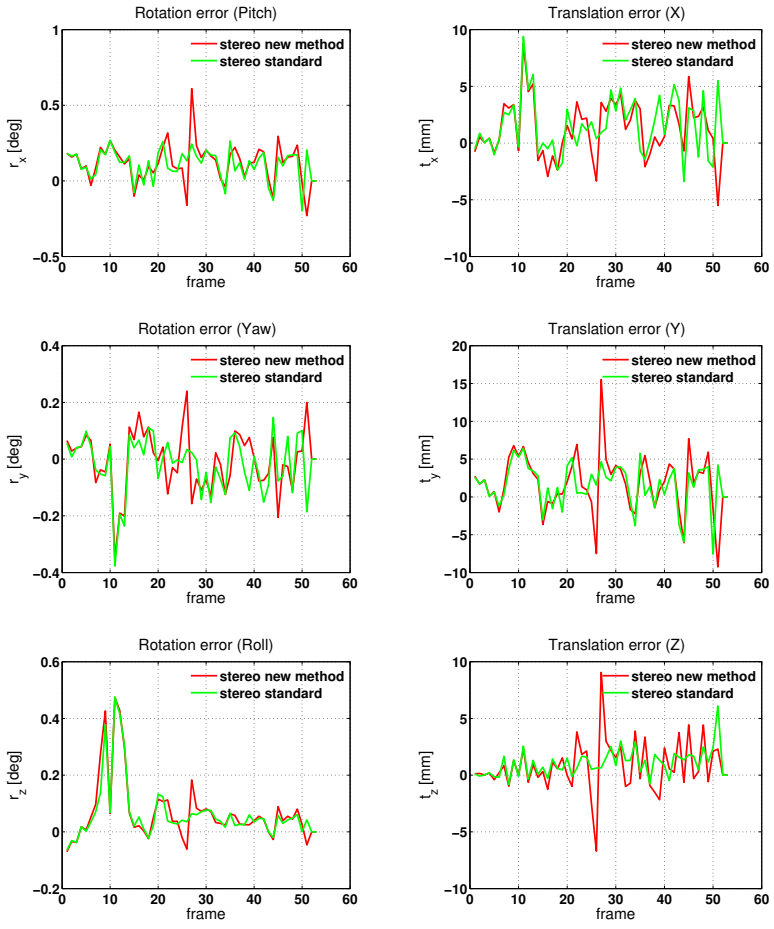


FIGURE 5.8: Comparison of the new stereo and standard stereo camera based methods with higher velocity, when the target spins at angular velocity of 4 deg/s and linear velocity of 20 mm/s and frontal to the camera.

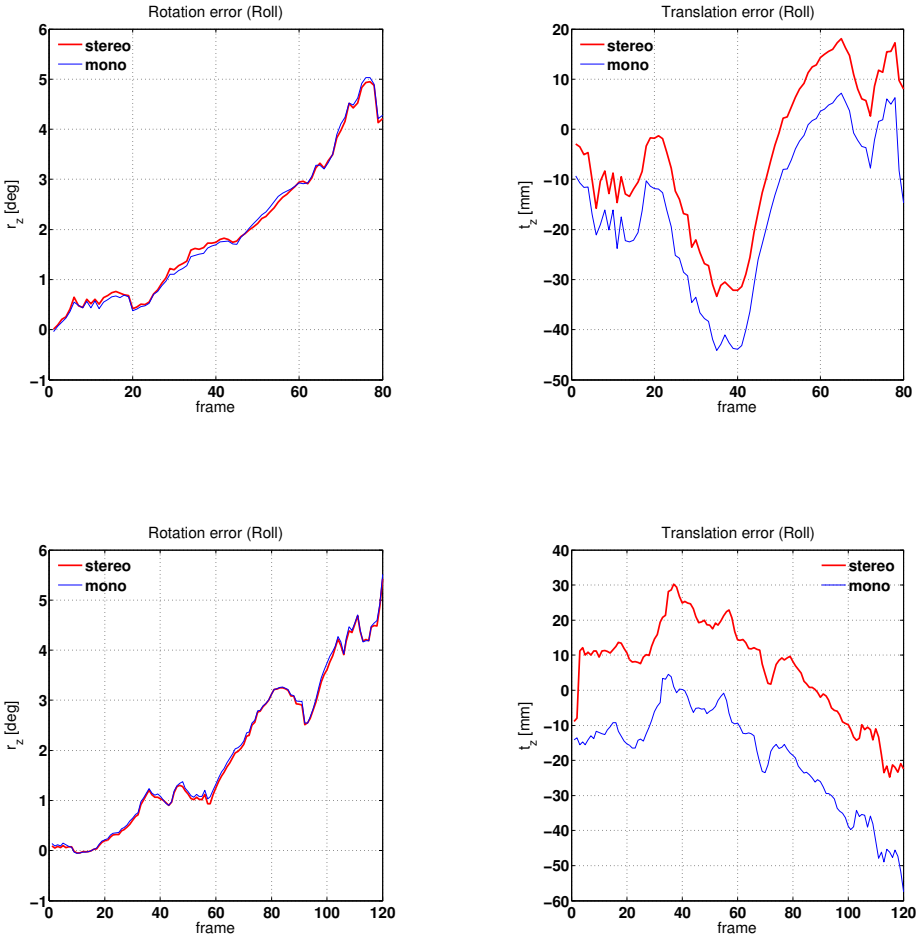


FIGURE 5.9: Drift of feature under various motion 2 deg/s and 10 mm/s. (Top row) roto-translation about spinning axis frontal to camera, and bottom the target nutated 5°

Chapter 6

Model-based Position Localization under Specular Reflection

In chapter 5, we have presented motion estimation and tracking in the absence of a geometric model (CAD model) of a client satellite, solely relying on image data and motion model. In this chapter, we present a model-based localization of the client satellite. Localization of a rigid-body in 3D space in the presence of the geometric model provides two major advantages. Firstly, it provides a priori information about the object for absolute pose estimation, enabling drift-free visual tracking. Secondly, the CAD model enables to exploit edge features for tracking which are robust to lighting, thanks to a model and an image edge matching along the edge normal (1D search). In this chapter, we address two configurations of a target satellite which occur in practice. First, we present position detection and tracking (position localization) of an attitude-controlled satellite (Section 6.1). The position localization here refers to estimation of the three translation parameters of a partially cooperative satellite in a sequence of images. In this configuration, the attitudes of servicer and client satellites are aligned by on-board sensors such as star tracker. The point here is that, the on-orbit servicing can be effectively performed just before the life span of the client (target) satellite comes to an end. Consequently, on-board attitude measurement sensors of the client satellite can be exploited. The second scenario will be presented in Chapter 7, which is orientation

and position (pose) tracking of a fully malfunctioning target satellite. The target satellite in this case tumbles freely about any of its axes.

6.1 Position Localization of a Partially Cooperative Satellite

In general satellites undergo translational and rotational motion in space. Under normal operating condition of a satellite, where Attitude and Orbit Control System (AOCS) are functional, the servicer and client satellites can be aligned with an aid of on-board attitude measurement sensors. In this scenario, position localization considers only translational motion. In fact, state of the art camera pose estimation methods are not robust for all object geometries and orientations (views). For instance, they have problems in case of object symmetry or frontal planar surface where in-depth rotations are ill-conditioned. Therefore, instead of tracking the full 6D motion parameters, we exploit the AOCS for alignment of the two satellites and localize position of the client with respect to the servicer. On the other hand, sensor fusion is well known in compensating drawbacks of measurement errors from sensors, however attitude measurement from AOCS is sufficiently accurate and robust in our particular application. Hence, the benefit of fusion of measurements from various sensors here is not feasible.

The main contribution of the thesis in this regard (based on our work [166, 167]) is, firstly it provides a model-based hierarchical position tracking method to handle change of image features along a range of distances, and localize an attitude controlled (partially cooperative) satellite. For this purpose, a robust histogram-based foreground and background segmentation is directly integrated into the position optimization loop in 3D space at mid-range, followed by an edge-based localization of the nozzle of the satellite at close-range. Furthermore, the performance of the classical edge-based tracking method is improved by integrating a simple motion model. Secondly, it provides a robust nozzle detection method, which adapts and integrates state of the art techniques which otherwise alone fail, addressing a real world challenging problem. Finally, the thesis contributes by providing realistic data sets similar to space scenario, and analyzes the tracking performance in terms of change of illumination, intensity, sun direction, motion and image compression.

6.1.1 Partially Cooperative Satellite

A partially cooperative satellite is defined here as a client satellite near its end of operational lifetime, but able to be controlled, hence on-board attitude measurement sensors are active and can be used to align with the attitude of the servicer satellite. Fully operational satellites maintain their orbit and are stable in space. This orbit control is achieved by providing fuel to maintain the orbit position, otherwise the external forces such as atmospheric drag (in LEO) and pull of the Moon and the Sun slow down, eventually lowering the orbit altitude. Similarly, the pointing of antennas is provided by attitude controllers and actuating devices such as reaction wheel, thruster or magnetic torquer. In contrast, when the satellite runs out of fuel or electronics fails, it becomes dysfunctional and inactive, corresponding to the term non-cooperative satellite. Also notice that a satellite can be a cooperative if markers are mounted on it to aid vision-based localization.

On-orbit servicing of a partially cooperative satellite provides several advantages. Firstly, the client satellite is active and one is able to communicate to assess the status of the satellite. Secondly, the active satellite means the on-board attitude measurement sensors are operational, i.e. an accurate on-board attitude measurement is possible, subsequently the relative attitude of the servicer satellite can be aligned. Thus, only position estimation and tracking is required to approach and manipulate the client satellite. In point of view of computer vision, this is equivalent to reducing pose estimation from 6 DOF to 3 DOF which enables to find search space efficiently during optimization.

6.1.2 Hierarchical Tracking and Detection

For the localization of the client satellite with respect to the servicer, we assume a partially cooperative client satellite, and the servicer is in a forced r-bar approach (the motion orthogonal to orbital velocity of the client) so that attitudes of the servicer and the client are aligned during the last 75 m. The on-board sensors of the spacecrafts provide necessary attitude measurements in this range. Here, we address only the most critical approach phase from 20 m. Beyond the range of forced r-bar approach, an autonomous rendezvous is generally feasible from a telemetry or a far-range camera, where measurements of only line of sight and range are required. Notice that the forced

r-bar rendezvous along the whole range of approach of the client is possible but consumes much fuel, therefore we restrict to the last tens of meters.

For a successful detection and tracking of the client, we employ two camera systems: a monocular and a stereo camera. The monocular camera is used for mid-range rendezvous (20 m to 5 m), with a larger focal length and smaller field of view, due to the higher distance to the target. On the other hand, stereo cameras are suitable for an efficient and robust localization at close range (from 5 m to 0.5 m), where a larger field of view and a significant disparity dramatically increase accuracy. During the approach, the localization system automatically switches between mid- and close-range.

We employ a hierarchical localization approach, by integrating robust detection and tracking methods because of the different image properties and camera system at each range. The algorithms in both ranges are based on the detection and tracking work-flow (Fig. 6.1):

- Detection block: beginning with the initial frame, or if the tracking fails at a certain frame, we perform a global search to find the client in the whole 3D viewing volume, at the given range. If detection has been successful, initialize the pose and go to local tracking, otherwise, report a failure.
- Tracking block: if a valid initial pose is available (either from the previous frame or from the detector), it is used to start a local search for the current frame. In case of failure, the detector will be called at the next frame.

However, internally they differ in the way the three main tasks (global search, local search and loss detection) are performed, concerning employed models, visual cues and search strategies. For example, loss detection at close range is performed by computing residual statistics, after matching contour points at the estimated pose and monitoring the percentage of outliers.

Good Visual Features for Localization

The performance of the detection and tracking algorithm in general depends on existing image features. For the localization of a satellite, we rely on contour and internal edges which are known to be robust to illumination variation. We exploit the silhouette

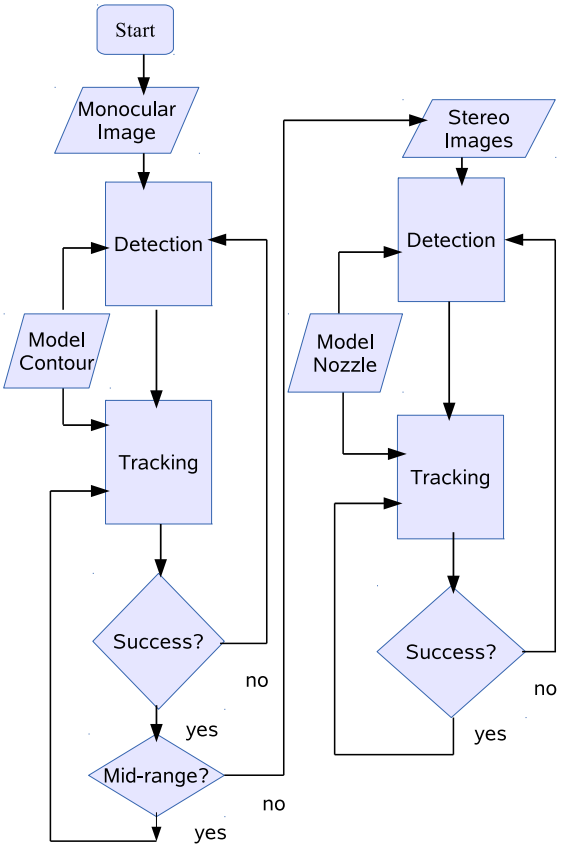


FIGURE 6.1: Hierarchical localization approach at mid- and close range. Both mid-range (left) and close range (right) localization consists of global detection and local tracking blocks.

contour (boundary) of the satellite within the field of view of the camera at sufficient range, otherwise the localization method relies on internal edges and primitive features.

In the case of an attitude-controlled satellite, only the position (translation parameters) should be estimated. The most suitable features for tracking are parts of the satellite which contain an engine thrust (nozzle), a launcher interface ring or a bracket (see Fig. 6.2). One of our goals here is to develop a tracking method that exploits the most common satellite features. In this regard, the silhouette contour of the satellite and thrust nozzle are readily obtainable features of most satellites. Therefore, we use the outer contour and nozzle of the satellite for mid-range and close-range tracking respectively. In fact, the nozzle of the satellite in both ranges is in a frontal view, despite slight attitude deviation, because of the attitude alignment of the servicer and the client satellites. We remark that the interface ring (the bigger circle) in Fig. 6.2 is a non-generic feature to track, as it may not exist universally in all satellites, hence we

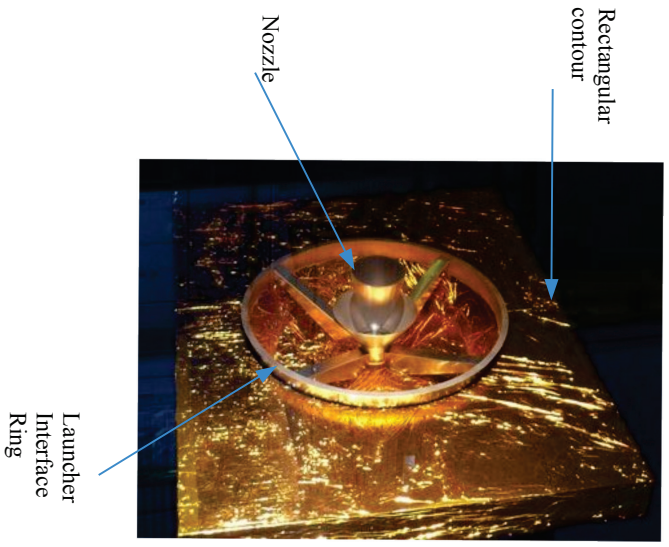


FIGURE 6.2: Features for satellite localization. The boundary contour and the nozzle are the most common elements of most satellites, and can be considered general features for visual tracking.

restrict to track the engine nozzle of the satellite.

6.2 Mid-range Localization

The localization of the contour of the satellite starts at about 20 m because of limited space in our facility. We employ a monocular camera with a small field of view, in order to increase the object resolution and allow a sufficient accuracy for model-based matching. In contrast, at such distances stereo cameras would require a large baseline and consequently a large field of view, resulting in a less accurate system due to the limited resolution. Moreover, such a setup would be very sensitive to calibration errors.

However, in order to cover the full range (down to 5 m) the field of view cannot be reduced too much, and therefore even a monocular system cannot reliably show some features on the surface, such as the launcher interface ring. Therefore, in this mid-range we rely on the outer satellite shape (silhouette contour), that can be identified as a bright region against a dark background, thus very suitable to a foreground segmentation approach. However, the boundary between the foreground and background may not be well separated due to shadow and poor lighting.

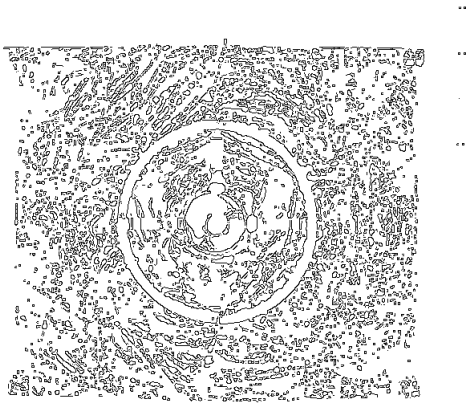


FIGURE 6.3: Edge-map at a distance of 6 m. The contours of the satellite are sparse and difficult for edge-based tracking, instead foreground and background statistics provide strong cues than contour edges.

The result of the position estimation cannot be expected to be as accurate as the close-range approach, however it will be certainly more robust and accurate than an edge-based approach. The whole satellite region offers a strong contrast against the background because of the specular nature of the multilayer insulation (MLI). In contrast, the reflection of MLI causes the nozzle to be almost undetectable in this range, as it can be observed from the related edge map in Fig. 6.3.

During this approach phase the required absolute accuracy is not as high as during the close-range approach: in fact, the main system requirement is to keep, at any range, a *relative* accuracy of in-depth translation of 1%. Therefore, at 20 m an error of less than 20 cm can be tolerated.

6.2.1 Contour Detection

We assume the client satellite is approximately known beyond the current mid-range through a far-range localization method, which is beyond the scope of the thesis. In order to find the silhouette contour of the satellite, we employ a brute-force global search on a large but a regular (x, y, z) grid covering the desired range to estimate the position (the Algorithm 1). The estimated translation parameters are then locally refined using the Nelder-Mead simplex algorithm [168].

Algorithm 1 Global Detection

1: Initialize the bounding box with minimum and maximum range in 3 dimension
2: Divide the bounding box into n number of grids in all the dimensions
3: **for** each grid n **do**
4: Weight minimum and maximum range according to the divided grid
5: Project the 3D parameters of the rectangle to 2D
6: Evaluate histogram-based segmentation and return cost
7: **if** *cost* $<$ *minimum cost* **then**
8: Return best estimate of 2D rectangle parameters that result minimum cost function evaluation
9: Back-project the 2D best estimate and compute 3D coordinates of the center of the rectangle

6.2.2 Contour Tracking

Since the geometric model of the client satellite exists, instead of a generic foreground and background segmentation a more effective model-based matching can be performed. In order to estimate the 3 translational degrees of freedom, the shape at any pose hypothesis (in this case translation and scale) is projected onto an image plane, collecting interior and exterior image statistics from the respective regions, and computing the *divergence* between statistics to be maximized (see Fig. 6.4).

In particular, at position $T = (T_x, T_y, T_z)$ the corresponding foreground and background regions $f(T), b(T)$ are computed, and a gray-value statistics using histograms $P_f^h(T), P_b^h(T), h = 1, \ldots, N_c$, where N_c is the number of histogram cells are collected. Here we use a power of two $N_c = 2^n$, typically $N_c = 64$, in order to quickly compute the histogram cell, $h = \lfloor (g/2^8)N_c \rfloor = g \gg (8 - n)$, where $g \in [0, 255]$ is any integer gray value.

Then, the Bhattacharyya divergence coefficient between the densities, defined by

$$B(f(T), b(T)) = \sum_{h=1}^{N_c} \sqrt{P_f^h(T) P_b^h(T)} \tag{6.1}$$

is evaluated, where $B(T) \in [0, 1]$ has to be minimized w.r.t. T, in order to enforce foreground and background separation. To keep a better shape for the objective function, rather $\log B(T)$ is maximized. In absence of derivatives, and in a few dimensions, local optimization is effectively performed by a direct search method, using the Nelder-Mead simplex algorithm [168]. Although the model used for tracking is the outer boundary of the object, the required output always refers to the position of the nozzle. In fact, since

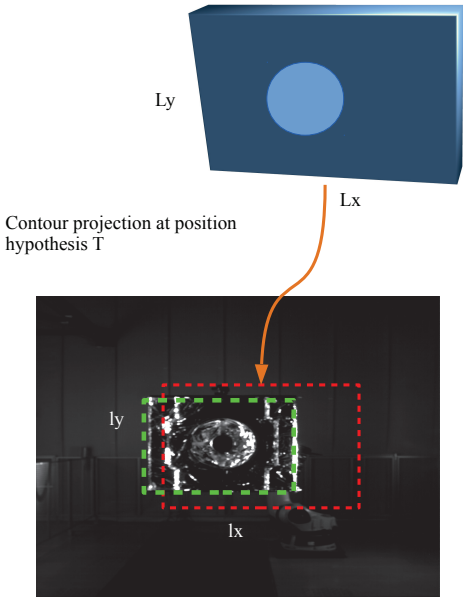


FIGURE 6.4: Perspective projection of a contour of a satellite at a given position hypothesis. Divergence of histogram statistics between foreground and background is maximized. After a few iteration, the re-projection of the rectangle fits to the image rectangle (green)

the relative location of the latter on the client body is known, the constant translation vector is simply added to the estimated position of the satellite contour.

6.3 Close-range Localization

The hand-over between mid- and close-range localization approaches takes place at about 5m distance. The mid-range tracking automatically switches to the close-range localization, that looks directly for the satellite nozzle, modeled as a 3D circular feature. Here, a stereo camera is employed for both detection and tracking, providing more accurate and robust measurements. The detection and tracking of the nozzle are based on edge extraction, where both location and orientation of edge pixels are determined through Canny detector [149] and Sobel filter respectively.

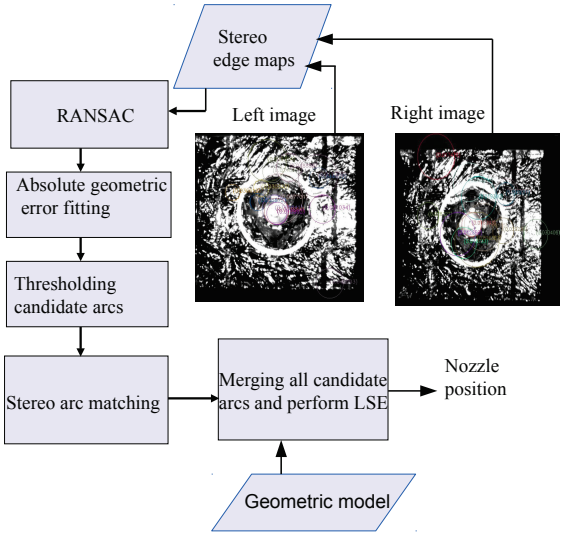


FIGURE 6.5: Nozzle detection of an attitude controlled satellite.

6.3.1 Nozzle Detection

After the hand-over between the two localization modules, the nozzle tracker is initialized by using the last pose estimated by the mid-range tracker. However, for more safety also in this case we need a general target detector, able to initialize the local nozzle tracker in case of failure. The detection module is thus responsible for a global recognition of the client nozzle in the incoming images. Therefore, localization robustness against clutter and noise, as well as sufficient accuracy, are vital in order to successfully initialize the tracker.

The client satellite produces strong background clutter (region outside nozzle rim), due to a specular reflection of the glossy and irregular surface under direct sunlight. The cluttered edge map (see Fig. 6.3) poses challenges to detection as well as tracking. We assume a frontal position and rectified stereo images, while both assumptions are relaxed during tracking, that work with arbitrary cameras and nozzle attitudes, and perform a robust bottom-up processing as depicted in Fig. 6.5. The bottom-up processing includes mainly edge detection, linking, circle detection, stereo matching, triangulation of the center and least square error (LSE) refinement.

Binary edge maps are first obtained by the Canny detector [149]. Subsequently, edge linking [169] is performed to recover connected, 1-pixel wide lines, which are broken at

detected junction points where two or more lines intersect, obtaining a set of labeled contours. This procedure may be computationally intensive, especially in a cluttered scenario, however it provides a reliable information for shape detection. Each connected line is then tested for circularity, that is, how well it fits an arc of a circle. The standard LSE fitting of a circle (L_2 norm),

$$LSE : \arg\min_{x_c,y_c,r} \sum_i \left(\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} - r \right)^2 \tag{6.2}$$

is not robust against outliers, that in our case are given by false edges, erroneously linked to the correct ones. A better strategy considers instead the L_1 norm of the geometric re-projection error

$$AGE : \arg\min_{x_c,y_c,r} \sum_i \left| \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} - r \right| \tag{6.3}$$

where (x_c, y_c) and r are the center and radius of the circle respectively, and (x_i, y_i) are the measurements.

Despite generally a higher difficulty of L_1 (absolute geometric error) optimization, in this case the result can be obtained efficiently through a recently developed algorithm [170], by partitioning the points into three sets: outside points, inside points and points on the circumference. This method performs at best when inliers have a low measurement noise, which is the case of the Canny detector that, at least in relatively low-noise images, shows a pixel accuracy.

The circle detection rate is further enhanced by first applying a RANSAC strategy (3 point circle fitting) for outlier detection by thresholding against a reasonable percentage, e.g. 30%, before performing the L_1 (AGE) optimization. Although this may increase the computational cost, especially in case of high clutter, in practice the RANSAC takes very few iterations to converge, again because of the low inlier noise, thus resulting in a negligible additional cost. A clear advantage of RANSAC is the fact that it explicitly selects inliers, resulting not only in an improved robustness, but also allowing to rule out non-circles, by thresholding the percentage of outliers.

After fitting, candidate arcs are pruned against a minimum spanned angle of 45 deg, and against an allowed range of radii, that are computed from the range of observable nozzle depths. That means, the radius of the 3D circle of the nozzle is re-projected onto

the image plane, using known maximum and minimum distance of the tracking range. Thus, the allowed range of the radii r_{min} and r_{max} are respectively,

$$r_{min} = f\frac{R_c}{Z_{max}} \tag{6.4}$$

$$r_{max} = f\frac{R_c}{Z_{min}} \tag{6.5}$$

where R_c, radius of the 3D circle, f focal length, Z_{min} and Z_{max} are closest and furthest distance of the nozzle from the camera respectively. The remaining circles are matched pair-wise between stereo images, by further setting thresholds determined by stereo constraints to exclude spurious matches. In particular, for a given nozzle radius and camera parameters, we test against the epipolar constraint between the two centers (approximately same y), the allowed disparity range corresponding to observable depths, the similarity of estimated radii as well as the known radius of the nozzle.

After the pruning, either none (in case of missing detection) or very few candidate circles will remain, each one leading to a hypothetical position, first computed by triangulation of the image centers. Those hypotheses often cluster together around the correct position, due to the fact that the detected nozzle edge splits into multiple arcs. Therefore, in order to obtain a more robust and accurate position estimation, we merge all candidate arcs (on both images), and perform a final nonlinear LSE, by minimizing the overall re-projection error, initialized by the average of triangulated centers. This method is in general robust to background clutter and illumination variation.

6.3.2 Nozzle Tracking

Assuming the nozzle position has been initialized with sufficient precision, the local tracking refines and updates the estimate in real-time, still relying on the Canny edge map, but this time minimizing geometric re-projection errors, after sampling the model into a set of points, to be projected onto both images. In contrast, extracting a circle under strong specular reflection and shadow is not robust enough for tracking, therefore instead of circle to point matching, we adapt the general point to point matching, without looking for special geometric features. This procedure is faster for real-time purposes and also more accurate and robust, because it is a local nonlinear least-squares estimation that minimizes a re-projection error, unlike a global, bottom-up search. Moreover, it

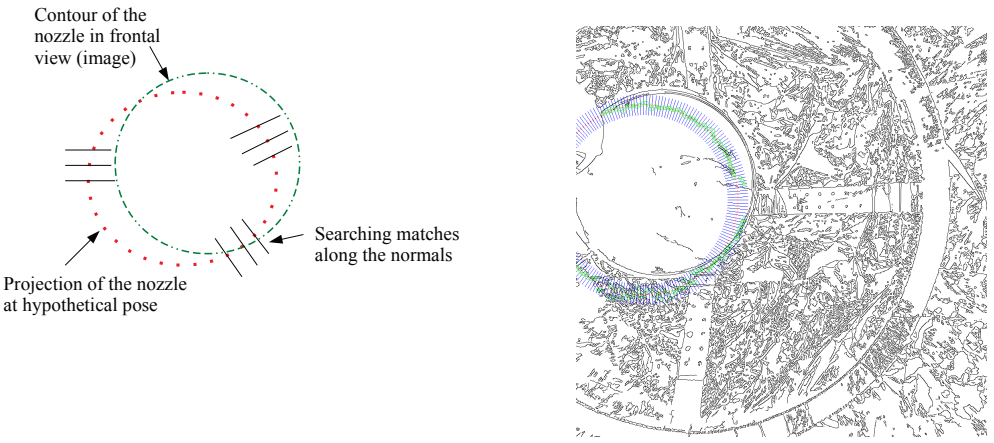


FIGURE 6.6: Re-projected nozzle contour at a given pose hypothesis in frontal view and contour of the nozzle in the image. (Left) the image and model edges (circle) are uniformly sampled for matching along the normal of the edge at each sample point. (Right) the nozzle contour re-projected at a given pose hypothesis in frontal view and edge map under good illumination. The optimization is a 1D search along the edge normal at each sample point.

does not anymore rely on the assumption of parallel cameras, thus avoiding the need for image rectification.

For the local tracking, model points at a given pose hypothesis (pose at last frame) are re-projected and associated to candidate edge pixels, which are determined by the Bresenham line drawing algorithm by sampling the Canny edge map, along the respective normals (as illustrated in Fig. 6.6), up to a constant maximum distance (validation gate), while also considering the corresponding edge directions as measured by the Sobel filter, that should be similar up to a 180° ambiguity, within a reasonable threshold of about ±15°. In Section 7.1, we present an adaptive validation gate which enables to handle apparently small targets in the image at relatively far distance.

As a comparison, it could be noticed here that the two most well-known model-based tracking methods [4] and [15], although both based on edges, perform pose estimation in rather different ways. The former samples 3D contour points and projects each point to the image in order to compute and minimize re-projection error. The latter samples model contours in 2D, after projecting lines on the image plane, and looks for point-wise edge correspondences between frames, similar to optical flow, in this context, referred to as "moving edges" approach, followed by pose estimation from 2D-2D point correspondences.

Despite similar performances under normal illumination condition, the two approaches may have different advantages and disadvantages, that have been explored in some detail in [5]. We adapt the former approach because of a more direct and accurate formulation, in that it minimizes directly the re-projection error from 3D to 2D. It is known that edges do not maintain a stable contrast over subsequent frames due to the glossy and specular surface properties on the boundary of the nozzle rim. In this regard, the former approach is more robust to this local illumination change than the moving edge approach.

The local tracking relies on the maximum-likelihood solution provided by the edge fitting, starting from the previously estimated pose [4, 59]. By assuming a Gaussian noise statistics for detected edges, we have a nonlinear least-squares problem, that can be solved by Gauss-Newton or Levenberg Marquardt optimization, improved for robustness by using an M-estimator (Section 6.3.2). Although we mainly deal with translational motion, let us first consider the more general setting of rigid body motion, given by the Euclidean group $SE(3)$ based on Lie algebra ([60])

$$T_g = \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \tag{6.6}$$

where R is a (3×3) rotation matrix and $\mathbf{t}$ a translation vector. Unlike Euler angles $R(\phi, \theta, \alpha)$ representation (see B.1 in Appendix B), a singularity-free parametrization (at least around the current estimate of T_g) is obtained by taking the tangent space $SE(3)$ to the manifold at T_{t-1}, given by an arbitrary vector μ_t

$$T_t = T_{t-1}\delta T(\mu_t) \tag{6.7}$$

where the local incremental transform is a singularity-free around $\mu = 0$.

More concisely, $\mu = (\omega, \mathbf{v})$ represents linear and angular velocity in local coordinates of T_{t-1}, and defines through the exponential map (see [60] for further details),

$$\sigma T = \exp\left(\sum_{i=}^{6} G_i \mu_i\right) \tag{6.8}$$

where G_i are 4×4 basis generators. Each 3D model point $\mathbf{x} = [x\ y\ z\ 1]^T$, sampled along the circular nozzle rim, is expressed in homogeneous coordinates and projected onto a

given camera $\mathbf{y} = [u\ v]^T$ by computing

$$\mathbf{y} = \pi(K \cdot T \cdot \delta T(\mu) \cdot \mathbf{x}) \tag{6.9}$$

where the operator $\pi()$ transforms from homogeneous to Euclidean 2D coordinates under perspective camera model and K is a projection matrix, obtained through camera calibration.

For pose estimation, we seek the minimizer of the cost function

$$\hat{\mu} = \arg\min_{\mu} \sum_{i}^{n_p} ||\mathbf{e}_i||^2 \tag{6.10}$$

$$= \arg\min_{\mu} \sum_{i}^{n_p} ||\mathbf{s}_i - \mathbf{y}_i(\mu)||^2 \tag{6.11}$$

where $\mathbf{e}$ is the residual, $\mathbf{s}$ is the 2D image coordinates of the matching edge to the projected point $\mathbf{y}$, and n_p is the number of matching pairs. This size may reduce, if some of the projected points do not have a correspondence within the search range, or they fall out of screen in one or both cameras. These points are counted as outliers: they do not influence the partial derivatives of residual (Jacobian J), however they are counted as a mismatch with maximum residual in the above cost function.

By minimizing the linearized residuals along the contour normals, at each Gauss-Newton iteration local pose parameters μ,

$$\mu = \left(J^T J\right)^{-1} J^T \mathbf{e} \tag{6.12}$$

are estimated and the T_{t-1} matrix is updated by Equation (6.7), until the increment μ becomes sufficiently small or a maximum number of iterations is reached. Upon convergence the final norm of residuals and the estimated error covariance $\left(J^T J\right)^{-1}$ provide a useful metric to detect a tracking failure. Since our method is suitable for monocular and multi-camera system, it can work also in case of failure of one camera, albeit a reduced accuracy and robustness compared to stereo camera tracking.

The above general formulation can also be applied to cases with reduced degrees of freedom, for example in presence of geometric symmetries. This is easy to accomplish

because of the local (object-related) parametrization: for example, since the axial rotation of a circle cannot be estimated, we can simply suppress the Jacobian column related to the respective generator G_i, and set $\mu_i = 0$ during Gauss-Newton iterations. As the attitude of the client satellite is aligned with that of the servicer using on-board attitude measurement sensors (e.g. star tracker and sun sensor), the formulation for the pose tracking given above reduces to position tracking by fixing the rotational part R to a known value and keeping only the last three generators (basis vectors) associated to translation in Equation (6.8).

Robust Tracking using M-estimator

As it can be observed in Fig. 6.6, several background clutters because of specular reflection result in false positives on the edge map that may be dense and very close to the nozzle edge. Robustness to such outliers is critical for the local tracking as well as detection. However, the latter case is even worse because local matching is performed independently point-wise, without the reduced ambiguity provided by continuous edge linking (Sec. 6.3.1).

Also in this context, we could apply a RANSAC strategy under a stereo re-projection model. However, for real-time efficiency we rather choose an M-estimator as adopted in the context of 2D vision-based control (visual servoing) [171], that adaptively thresholds outliers based on robust median statistics, while re-weighting inliers according to the Tukey bi-weight function

$$w_i = \left(b^2 - \left(\frac{e_i - \bar{e}}{\sigma} \right)^2 \right)^2 \tag{6.13}$$

and $w_i = 0$ for $|e_i - \bar{e}| > b\sigma$. Here, $b = 4.6851$ is the standard threshold for Tukey's function, and the scale σ is the standard deviation of inlier data, estimated using the median of absolute deviations (MAD) from the median residual $\bar{e}$, and used to compute the inlier weights. The Gauss-Newton update in Equation (6.12) is then replaced by

$$\hat{\mu} = \left(J^T W J \right)^{-1} J^T W \mathbf{e} \tag{6.14}$$

where W is a diagonal matrix containing the weights.

Temporal Constraint for Motion Prediction

In spite of introducing the robust estimator into a maximum-likelihood solution to edge fitting, it is likely to lose track because of a higher number of outliers than inliers. The procedure can be easily improved for robustness by using a Kalman filter with a simple motion a priori. This is vital, although the global detector can provide again the absolute pose when local tracking fails, this is not always the case: low-contrast edges, severe background clutter and shadows may cause a long sequence of failures.

To address this problem, we adopt a simple motion model to predict the client satellite pose based on Kalman filtering. To this end, we assume the client satellite undergoes a constant translation velocity $\mathbf{v}$ and a constant rotation velocity ω between two camera frames. Notice that the satellite attitude is controlled and the trajectory is smooth, therefore the motion can be modeled with a constant velocity and small uncertainty.

Let $\mu = [\Omega^T \ \mathbf{t}^T \ \mathbf{v}^T \ \omega^T]^T$ be the state of the system, where $\Omega \in \Re^3$ is the rotation vector associated to the rotation matrix R, then

The state dynamics of the satellite can be described by a non-linear motion model,

$$\begin{aligned} \Omega_{k+1} &= \log_{SO(3)} e^{\hat{\omega}_k} e^{\hat{\Omega}_k} \\ \mathbf{t}_{k+1} &= e^{\hat{\omega}_k} \mathbf{t}_k + \mathbf{v}_k \\ \omega_{k+1} &= \omega_k + \alpha_k \\ \mathbf{v}_{k+1} &= \mathbf{v}_k + \mathbf{a}_k \end{aligned} \tag{6.15}$$

where $\hat{\omega}$ and $\hat{\Omega}$ are skew-symmetric matrices corresponding to the respective vectors, $\log_{SO(3)}$ is the inverse of rotation matrix exponential, that gives the rotation vector corresponding to rotation matrix R through the Rogrigues' formula.

The relative accelerations α_k and $\mathbf{a}_k$ between the servicer and the client are modeled as a Gaussian additive noise of zero-mean. The motion model used for state propagation in Equation 6.15 has been shown to be robust and stable [172]. In order to compute state update through an extended Kalman Filter, the above dynamics are linearized by differentiating the logarithmic function in SO(3). On the other hand, the measurement equation of the filter relies on pose estimates provided by the localization system described above. Hence, an identity measurement matrix is assumed for the position

part, $H = [I\ 0]$. As described in the previous Section, the client is well-controlled and attitude is aligned with that of the servicer with the aid of on-board sensors. Therefore, prediction concerns only the three translation parameters.

6.4 Evaluation

In the following the localization methods described in Sections 6.1 to 6.3 are experimentally evaluated using a satellite mock-up (Fig. 6.2). We employed a versatile Optical Rendezvous, Docking and Landing (RVDL) camera system, including an illumination subsystem, developed by Kayser-Threde GmbH shown in Fig. 6.7.

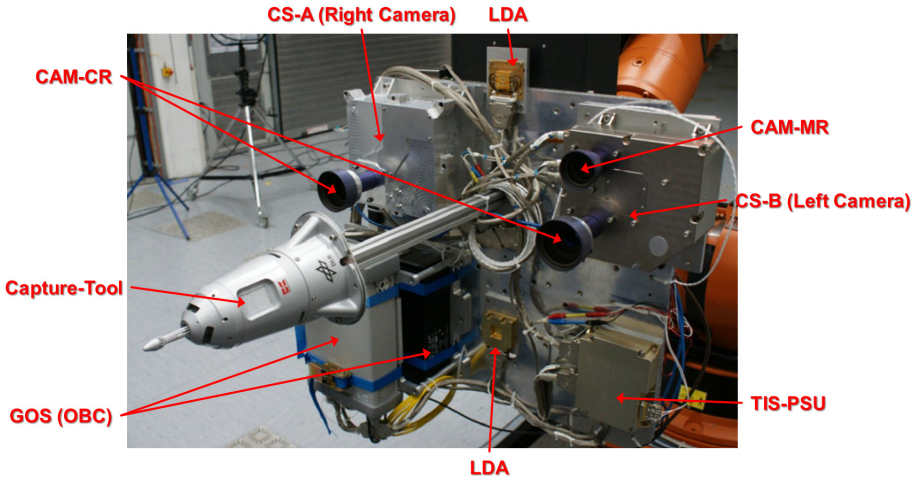


FIGURE 6.7: The VIBANASS equipment employed to evaluate the tracking system. Image courtesy of OHB System AG (former Kayser-Threde GmbH).

The equipment consists of stereo Camera Systems (CS), the Target Illumination System (TIS) with two illumination units (LDA) and a Ground Operation System (GOS) in flight configuration. The equipment contains three camera systems (a mid-range camera, and close-range stereo cameras), communication interfaces, hardware controller and a power supply module. At close-range, two CS are used as a stereo system with a 50 cm baseline. In addition to the hardware, the GOS comprises software to command the equipment, to receive and compress images, and finally to estimate the position of the target in real-time.

Fig. 6.8 shows a schematic diagram of the general test setup, comprising the target mock-up (left robot) and the Vision Based Navigation Sensor System (VIBANASS) equipment mounted on the right robot. This robot is mounted on a rail for the movement in the

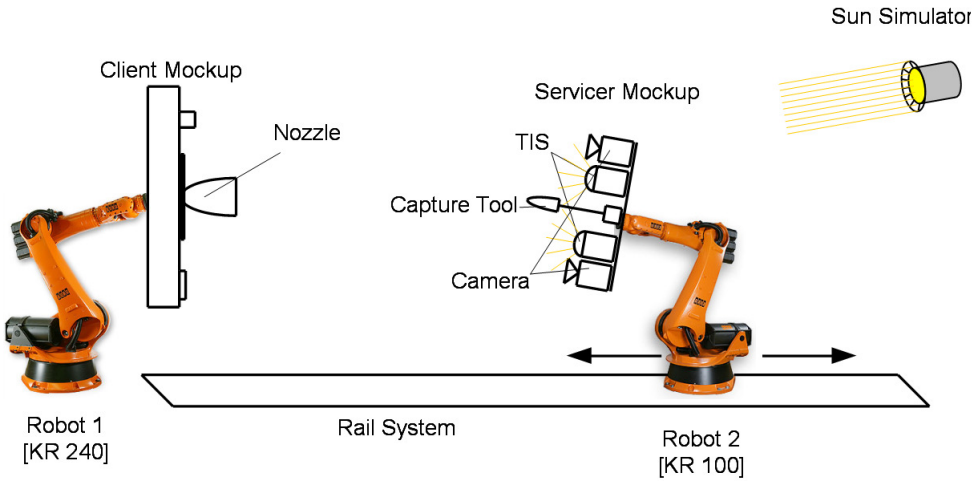


FIGURE 6.8: Schematic diagram of proximity operations test setup at EPOS facility (DLR). Two six degrees of freedom robots are employed to simulate servicer and client satellites. To achieve near micro-gravity and floating, a control strategy is used to compensate gravitational acceleration.

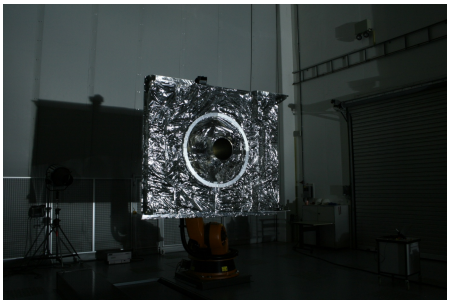


FIGURE 6.9: Mock-up of a full-scale client satellite at the EPOS facility. It consists of a thrust engine (nozzle) and a launcher interface ring (big circle).

line of sight direction. The sun simulator shown on the right is a 12 kW floodlight with a sun like spectrum and optical power.

The goal of the experiment is to evaluate the proposed methods to a real world problem under harsh illumination condition. To verify the functionality of the whole system, pre-recorded images are used as input for the localization algorithm and the estimated position of the client was compared to the measured position. The VIBANASS camera system is used to record monocular images at mid-range (>5 m) and stereo images at close-range (<5 m). At close-range, the client is additionally illuminated by the TIS. Results of the visual tracking algorithm are computed in different reference frames, located on the the left camera center at close-range, and on the monocular camera center at mid-range, but later transformed to same reference frame. Fig. 6.9 shows the target satellite mock-up, illuminated by the sun simulator.

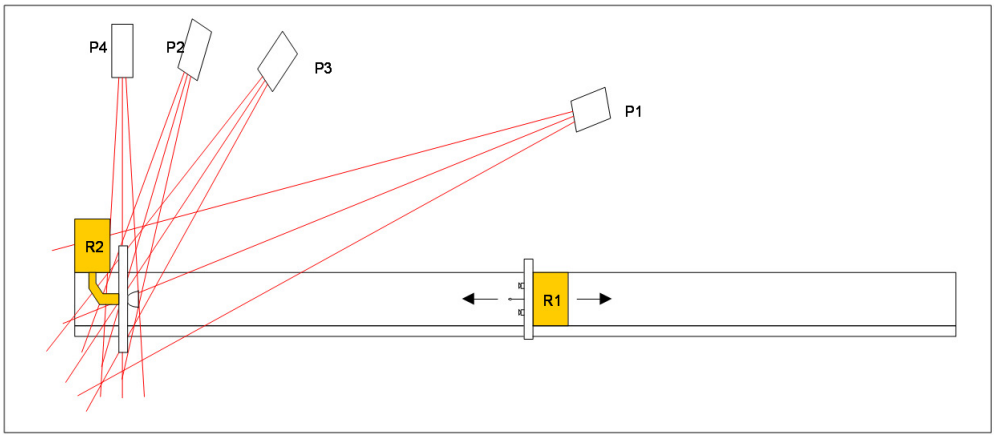


FIGURE 6.10: Positions of the sun simulator with respect to the target, measured by the angle between the camera line of sight and the sunlight direction. Image courtesy of OHB System AG (former Kayser-Threde GmbH).

FIGURE 6.11: Approach, as seen by a surveillance camera at a distance of 15 m (left) and 0.5 m (right).

A number of trajectories are tested, starting from a distance of 20 m and ending at a parking position of 0.5 m, measured between the camera and the target frame (center of the nozzle rim). Those trajectories cover a linear and non-linear motion (Table 6.1) and are used under various illumination conditions. Moreover, in order to assess the performance of the tracking method, the trajectories are evaluated with four sunlight directions. Fig. 6.10 shows four positions of the sun simulator. The positions of the sun simulator differ in the angle between the camera line of sight and the sun direction. In addition, a test case without sunlight (source of light 100% provided by TIS) is defined. In order to show a wide variety of environmental conditions several test cases are selected, with varying trajectory, sun position and intensity of an artificial source of light. Fig. 6.11 shows the approach trajectories at a distance of 15 m and 0.5 m, as seen by a surveillance camera. In the following, we present results and discussion of the mid- and close-range localization.

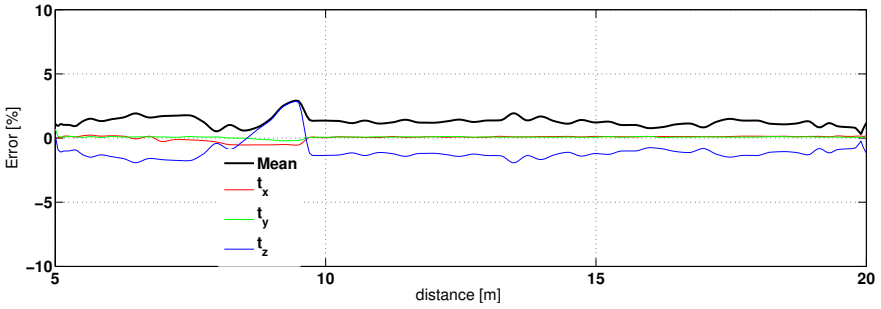


FIGURE 6.12: Mid-range tracking error. A servicer approaches a client at a nominal motion, approximately at a constant speed, at the sun direction of 90°.

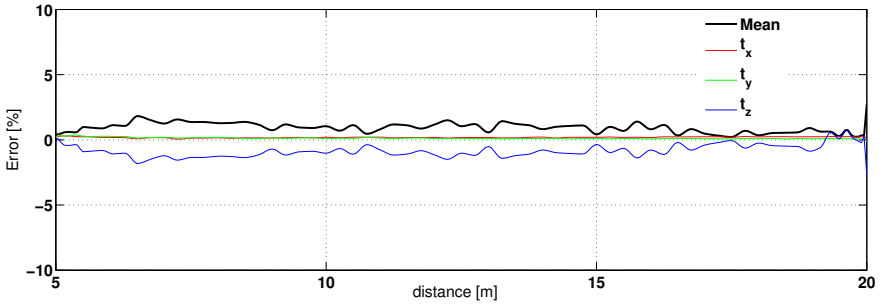


FIGURE 6.13: Mid-range tracking error at the sun direction 45°. Unlike the close-range tracking, the influence of the illumination at this angle is insignificant due to the effective foreground and background segmentation.

6.4.1 Mid-range Position Localization

A gray-scale image is acquired with a CMOS Camera sensor at a resolution of 1024×1024 pixels and a lens aperture of F-number 4 and focal length of 20.4 mm. Since the mid-range is approximately pre-determined from far range, detection is approximately semi-global, meaning the range of initial position is roughly known in the order of a few meters. Therefore, the detection algorithm can effectively search accurate position in a few iterations. For this purpose, the search box is defined initially as $-1 \text{ m} \leq T_x \leq 1 \text{ m}$, $-1 \text{ m} \leq T_y \leq 1 \text{ m}$, and $5 \text{ m} \leq T_z \leq 22 \text{ m}$. In fact, the position detection works beyond the given specification, but it becomes slower.

The mid-range tracking, which is model-based separation of histogram statistics between the object foreground and background is robust to poor illumination condition (see Fig. 6.15). Here we consider, a typical trajectory where a servicer approaches the client at a nominal motion profile, approximately constant speed. It appears that the change of the illumination direction in the mid-range does not affect the localization algorithm significantly due to the effective foreground and background segmentation. However,

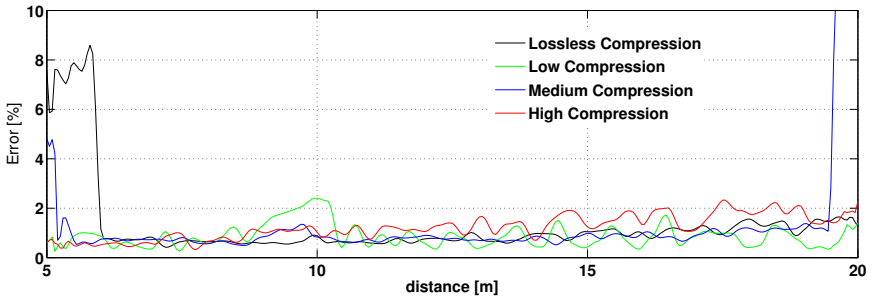


FIGURE 6.14: Effect of image compression on mid-range localization.

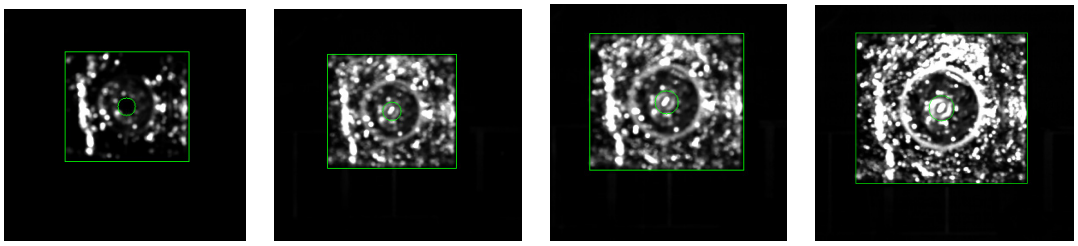


FIGURE 6.15: Model based foreground and background segmentation. The alignment of image contour with projected model contour (marked with green around the contour) at correctly estimated position. Images (left to right) are at positions 19m, 18m, 16m and 10m.

we observe a minor difference along the optical axis. This is, for example, evident in Fig. 6.13 where the client was tracked at the sun direction of 45° for the same trajectory in Fig. 6.12 at the sun angle of 90°.

On the other hand, the effect of quality of image in terms of compression, while resolution was maintained, on tracking in the mid-range was analyzed. The goal of this evaluation was to assess how accurately we can track compressed images. Notice that transmission of compressed images to ground enables to save bandwidth. As we can observe in Fig. 6.14, the compression did not affect the tracking. The evaluation includes lossy compression with a low, medium and high compression ratio and lossless compression. As it can be noticed in Fig. 6.14, the error suddenly increased at distances near to 5 m, when the switching occurred between mid- and close-range cameras. This effect was significant in the case of tracking with the lossless compressed images.

6.4.2 Close-range Position Localization

Gray-scale image sequences are acquired with a stereo CMOS Camera at a resolution of 1024×1024 pixels, and a lens of aperture with F-number 4 and focal length 6 mm. The

approach of the client satellite during the last 5 m are generally a critical phase, therefore various motion profiles at four illumination direction are evaluated. For this purpose, five types of motion trajectories (Table. 6.1) under various illumination conditions and sun directions are analyzed. In order to capture the real scenario, attitude disturbances acting on the servicer side, that cause oscillations of the camera (planar and pan-tilt rotation) with a given amplitude and frequency are introduced to the trajectories. It can be noticed that especially the pan-tilt rotation, although being small on the servicer side, introduce significant horizontal and vertical motion of the client nozzle in camera coordinates.

TABLE 6.1: Motion profiles used for evaluation of localization at close-range under various illumination condition

Motion Profile	Motion Description
1	Linear translation along Z
2	Linear translation along Z, plus a small rotation with sinusoidal oscillations about the Y axis
3	Linear translation and a small linear rotation about Z, plus linear translation disturbance and small rotation (the latter with sinusoidal oscillation) about Y
4	Station keeping, fixed at 5m distance, plus oscillating translation disturbance along Y and a small oscillating rotation about Y
5	Station keeping, fixed at 0.8m distance, plus the same disturbances of motion profile 4

In the following, the experimental results of the nozzle detection and tracking are discussed.

Nozzle Detection

In order to verify the accuracy and robustness of the nozzle detection method and compare with the state of the art, several trajectories and illumination direction were used. The performance evaluation is based on errors generated in 3D localization (accuracy), as well as percentage of correct detections (robustness) under conservative parameter settings. The detector parameters are important to remove undesired false positives. These include maximum disparity, the ratio of radii of the corresponding circles in the stereo images (ideally one), ratio of estimated radius to actual radius in 3D space (ideally one), the angle subtended by an arc of a circle (arc angle), and the threshold of

TABLE 6.2: Comparison of nozzle detection

Method	L1	L2	RANSAC	**RANSACL1**	RANSACL2
Correct detections	92	109	74	106	114
Total detections	139	147	87	114	165
Correct localization[%]	66.19	74.15	85.06	**92.98**	69.09

stereo re-projection error after merging all candidate arcs. To compare various detection approaches, we fix the same parameters for all fitting methods shown in Fig. 6.16. Particularly, the arc angle and threshold of re-projection error required a careful tuning to avoid false detection. In this work, a minimum arc angle of 45° and an overall re-projection error of 6 pixels were used for all the experiments.

The state of the art methods for the detection of nozzle in a frontal view (circle) are compared. The basic strategy of the detection is similar (Fig. 6.5), but differ in metric of the cost function, namely absolute error(L1), least square error (L2), RANSAC, hierarchical RANSAC and L1 (RANSACL1) and hierarchical RANSAC and L2 (RANSACL2). It is known that RANSAC is robust to outliers. Moreover, when it is combined with an other robust metric L1 (RANSACL1), the performance of localizing the nozzle of the satellite improves over all test sequences (see Fig. 6.16 and 6.17 as well as Table. 6.2). This is because of the complementary effect of the two robust methods (see Fig. 6.2). In particular, RANSAC performs poorly when there exist more outliers than inliers, in this case introducing L1 improves the detection performance. This combination is better in that it provides less position errors and a few false detections (false positive) consistently. It is worth noting in Fig. 6.16 that the RANSACL1 and RANSACL2 localized the nozzle position accurately (relatively a few errors), however the errors of RANSACL2 in the first five frames, are too high to initialize the local tracker (absolute errors more than 100 mm in Z-axis).

Nozzle Tracking

Based on the trajectories in Table 6.1 and associated ground truth, the nozzle tracking is evaluated and the influence of the following are analyzed:

- effect of the sun direction,
- effect of a Target Illumination System (TIS),
- effect of a compression artifact.

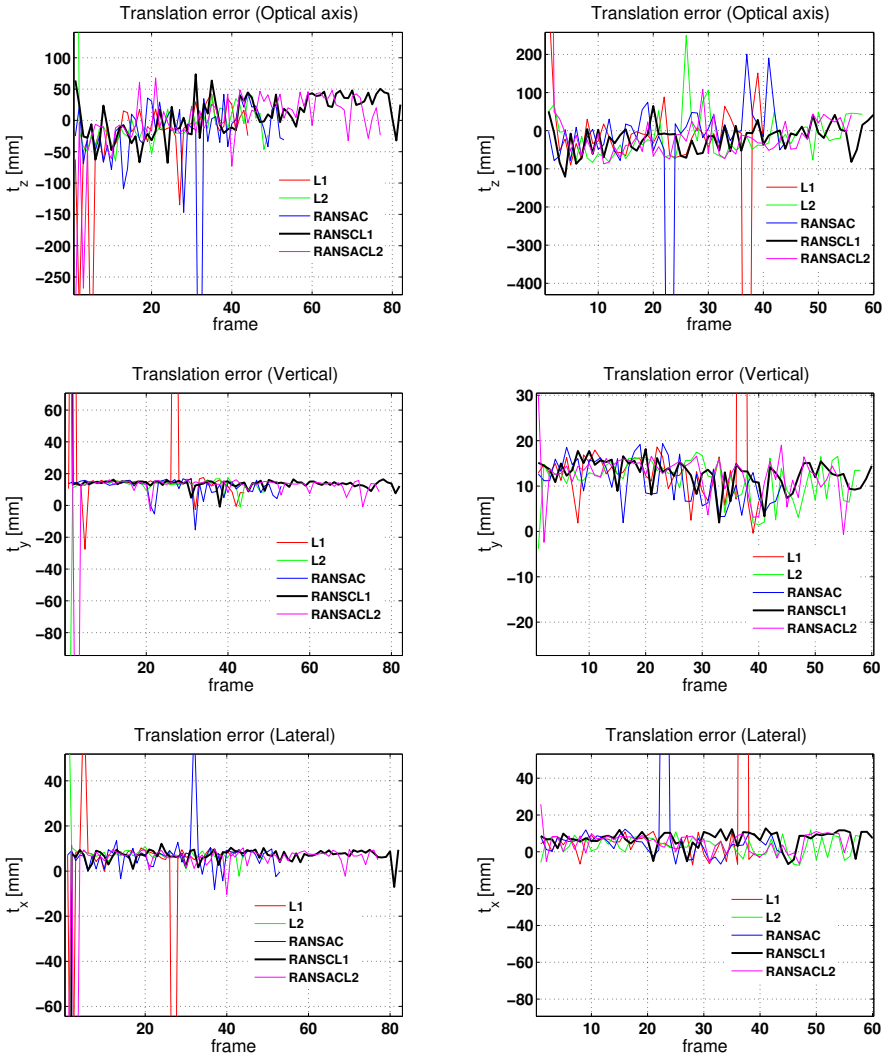


FIGURE 6.16: The translation error of the nozzle detection. The client satellite was illuminated 20% (left) and 100 % (right). The motion trajectory is linear along Z-axis (motion profile 1).

Effect of Sun Direction

The direction of the sun affects visual tracking. This effect is higher particularly for specular objects. Therefore, it is desirable to assess the tracking algorithm under certain critical illumination direction. It is impractical to evaluate under all possible illumination directions. However, typical direction from the on-orbit servicing mission can be retrieved. The orbit of the satellite may be pre-determined and can be detected from which range of direction the sun appears during on-orbit servicing. Therefore, we consider here four directions, which are obtained by dividing latitudinal space into $90°, 66°, 45°$ and $15°$. The direction of the sun on the client satellite (nozzle) in this

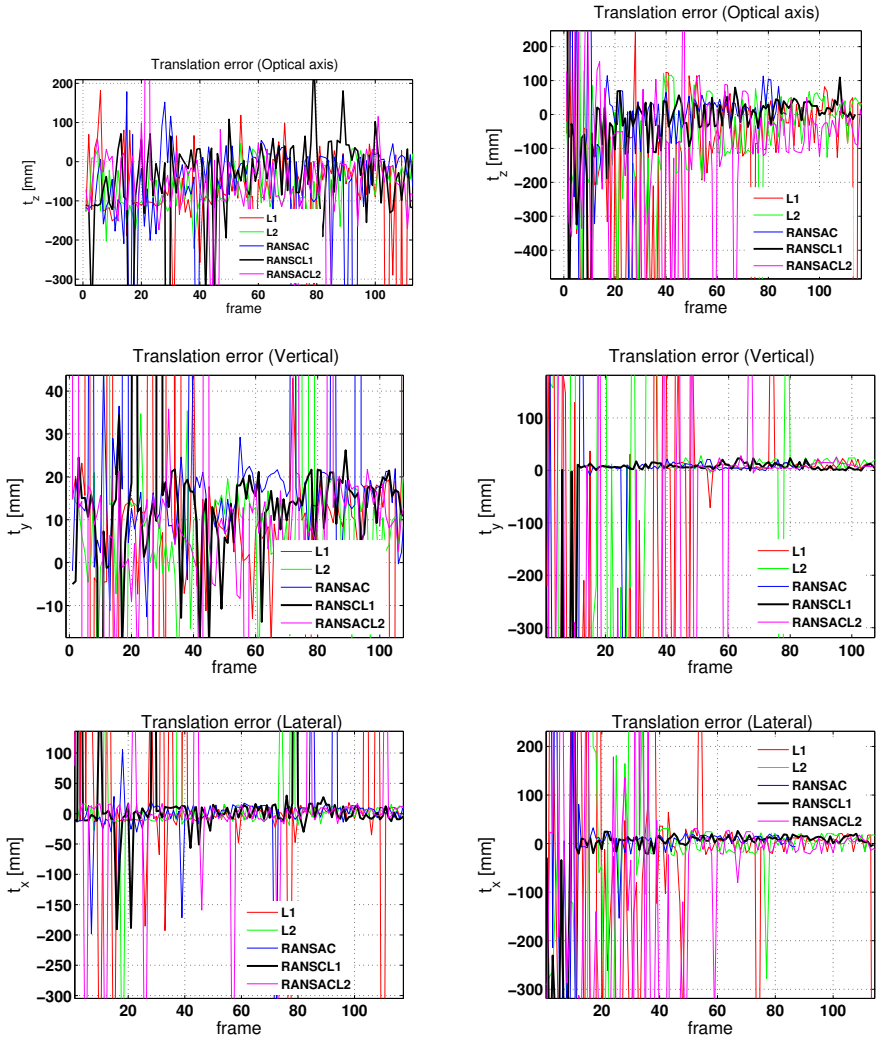


FIGURE 6.17: Nozzle detection. Translation errors for motion profile 2 (left column) and 3 (right column) at illumination level of 10% and 5% respectively.

experiment, particularly at close-range, varied from perpendicular to near parallel with respect to camera line of sight.

In order to perform a comparison at various sun directions, each motion profile was at the same illumination level of 10% (ratio of intensity of laser diode arrays (LDA) to that of the sunlight simulator). Accordingly, we compare effects of illumination for each motion profile. In motion profile 1 (pure translation along Z), as we can see in Fig. 6.18, the translation error at the sun direction of 90° is higher, particularly in the first few meters. This is evident for the dominant motion along the camera optical axis (Z). This is because, the sunlight at an angle of 90° was not sufficient to illuminate the interest region (nozzle), which is in a shadow in spite of an additional target illumination system. In fact, the level of the TIS could be increased to improve the tracking accuracy (for

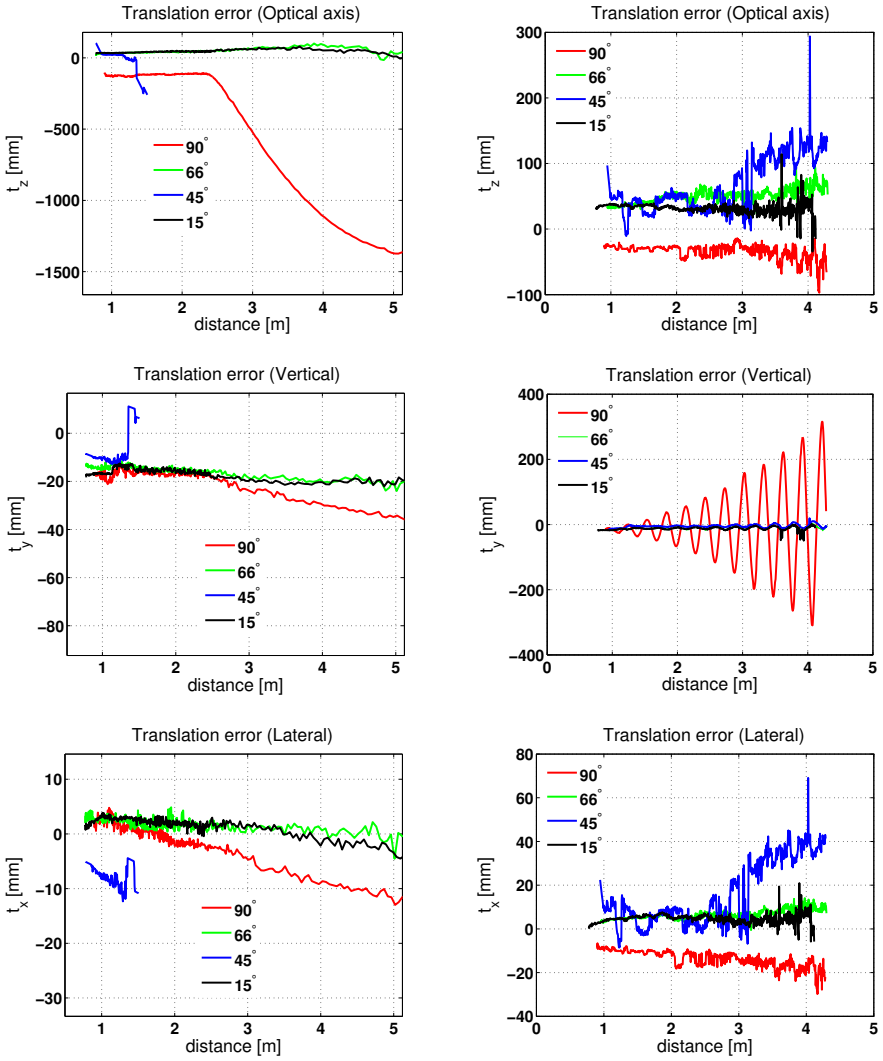


FIGURE 6.18: Effect of sun direction on estimation of position. Motion profile 1 (left) and motion profile 2 (right).

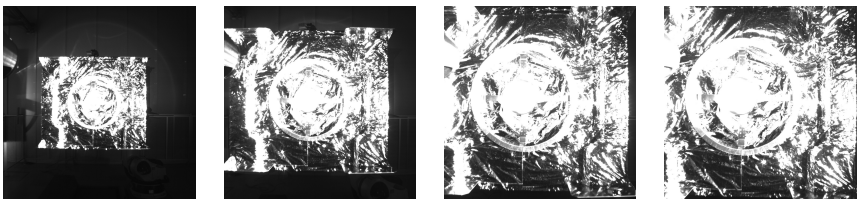


FIGURE 6.19: Effect of Target Illumination System. A 10% TIS saturates the image at a sunlight direction of 45°.

example by 20%, see Fig. 6.21). In contrast, at the sun direction of 45° and the same TIS level of 10% (Fig. 6.18), the images are completely over-saturated (see Fig. 6.19), consequently the tracking fails to track the position until the servicer approached to 1.5 m, in which case nozzle edges are distinctive to track despite over-saturation.

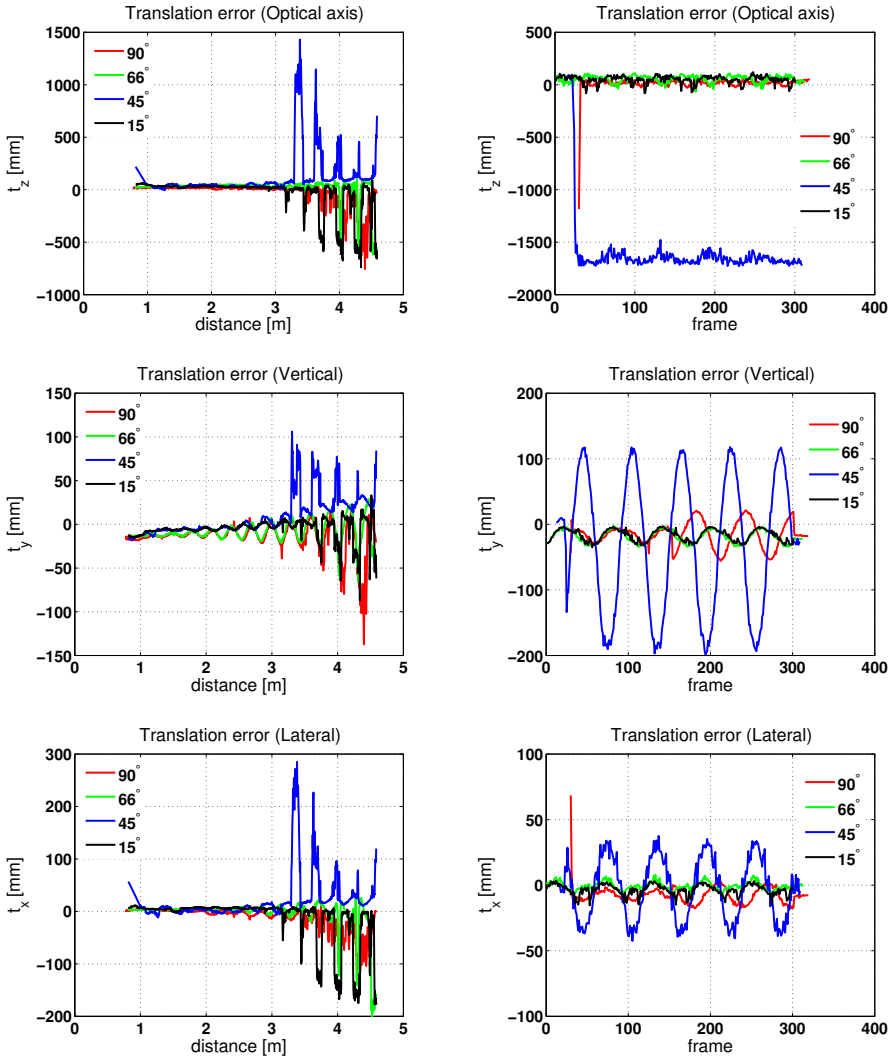


FIGURE 6.20: Effect of sun direction on estimation of position. Motion profile 3 (left) and motion profile 4 (right).

Similar observations related to poor illumination (90°) and over-saturation (45°) occurred also for the remaining of the motion profiles (Fig. 6.18 (right) and 6.20). However, in these cases the accuracy of the tracking improved at the sun direction of 90° and 45° because of the motion profiles which consist of lateral motion, that enable to change view point of the camera. In particular, the over-saturation at sun angle of 45° improved significantly for motion profiles which consist of lateral motions (translation and oscillation).

On the other hand, the tracking errors at the sun directions 66° and 15° are lower for all motion profiles. Notice that at the sun directions 90° and 45°, the tracking error is relatively high. Therefore, we expect that for a given TIS setting of 10%, the tracking performance improves at the sun direction close to 90° and 15°. The experimental result

is consistent with our expectation in that the error is lower at sun directions 66° and 15°.

Effect of Target Illumination

Depending on the direction of the sun to the region of interest of the satellite during on-orbit servicing, the target illumination system is essential particularly at a close-range approach for an accurate and robust tracking. TIS should provide as much light as necessary for poorly illuminated targets due to shadow. As described above, when the sun is perpendicular to the line of the sight of camera, it is very difficult to track a poorly illuminated client satellite. Therefore, we employed the Target Illumination System mounted on the servicer, which can be controlled from ground to change the level of illumination as necessary. In our experiment, the level of TIS is described relative to that of the sun simulator, which is 12 KW metal halide lamp with a spectrum near the real sun spectrum, i.e the percentage of the target illumination system to sunlight. Accordingly, we considered four levels of TIS setting: 5%, 10%, 20% and 100%. For example, 10% TIS represents that the LDA of TIS provides $\frac{1}{10}th$ of the sun illumination, and 100% TIS, lighting is fully provided by TIS (full shadow or eclipse condition). We evaluated the nozzle tracker at various levels of illumination and the sun direction of 90° for motion profile 1 and 3 (Fig. 6.21).

As it can be observed in Fig. 6.21 (left), a higher tracking error at the illumination level of 10%, is mainly due to shadow that is created by side illumination. The additional illumination provided by TIS seems to be not sufficient. On the other hand, the tracking accuracy improves significantly as the illumination level is increased (Fig. 6.21). When fully illuminated with TIS (100%), the performance of the tracker significantly increases. Thus, the effect of the shadow is minimized because of the TIS, consequently the tracking performance improved.

The motion profile influences the performance of the nozzle tracking under the same condition of illumination. For example, the tracking error for motion profile 3 (Fig. 6.21), which consists of oscillatory motion in two of the axes, all the TIS level result in similar tracking performance, at least during the last few meters. This is because, the oscillation of the client satellite promotes an appearance change as a result of specular reflection. However, as we can observe in Fig. 6.21 (right), at lower illumination level (5% and

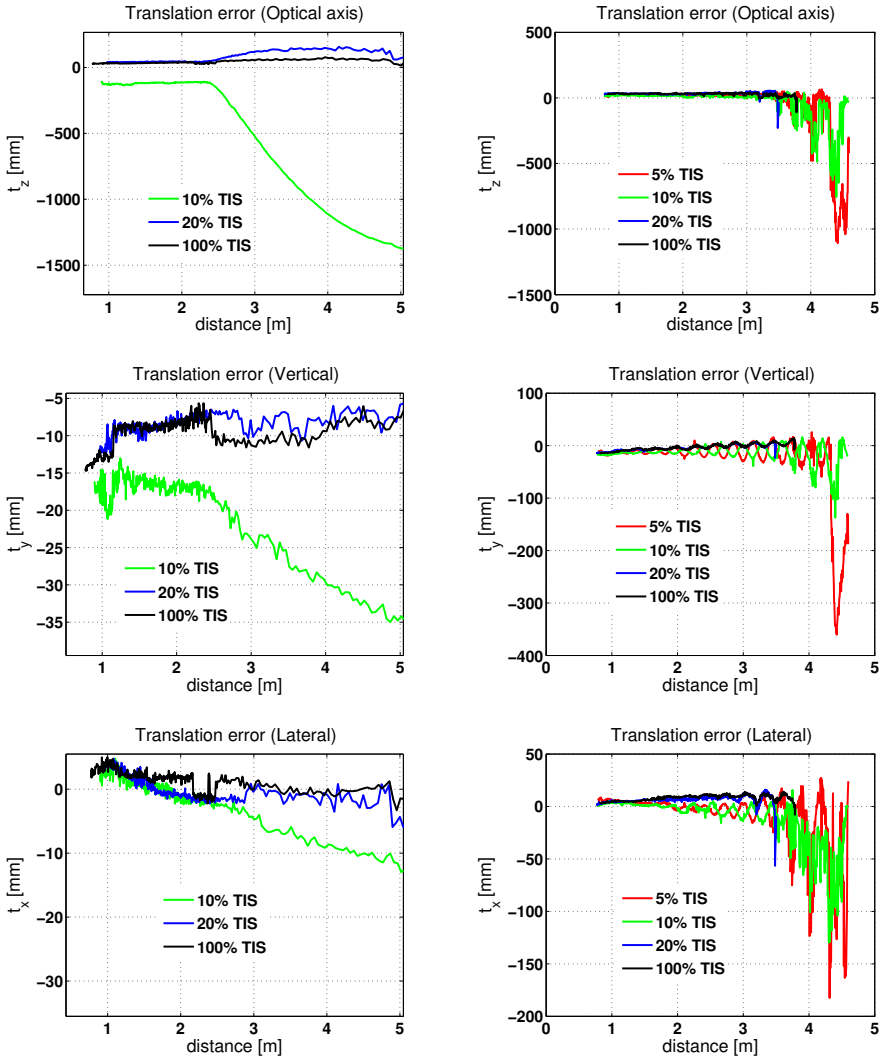


FIGURE 6.21: Effect of a Target Illumination System (TIS) on the tracking at the sun direction of 90°. Motion profile 1 (left) and motion profile 3 (right). The error is higher in the case of low level TIS.

10%) the tracking error is slightly higher as in the case of motion profile 1. This means, in general it seems that the intensity of the TIS should be higher irrespective to the motion profile.

Effect of Compression Artifacts

Some missions require sending image sequences down to ground for processing and re-send the processed higher level information such as 3D position of the satellite. In this case, compression of the images saves a large bandwidth, thus the impact of image compression settings on visual localization errors is a topic of great interest. Similar to

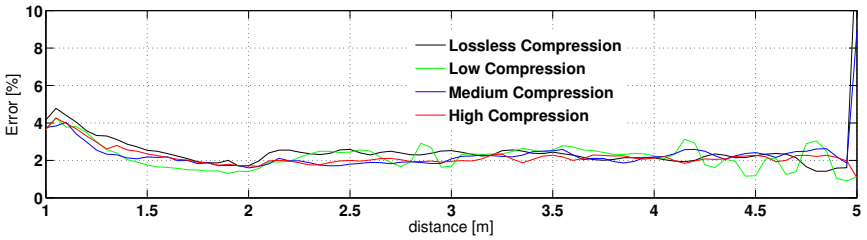


FIGURE 6.22: Effect of an image compression on close-range tracking. The error is comparable to the values achieved with good illumination.

section 6.4.1 discussed for mid-range tracking, we analyze the effect of various compression factors for the close-range approach. Several trajectories, under good illumination conditions, have been repeated with different compression settings. The resulting error has shown to be almost equal to the error achieved under good conditions, and appears to be independent of the compression rate. This is evident in Fig. 6.22, and it appears that the effect of image compression has less impact on accuracy of the tracking.

On the other hand, the worst case scenarios were handled by introducing a simple motion model to predict the position. When the nozzle contour is not discernible from the background, both detector and tracker fails. Therefore, an integrated motion model based on temporal constraint (constant velocity model) prevented the tracker from loss of tracking (see Fig. 6.23). The performance of the tracking under various illumination condition and image compression are shown in Table. 6.3. Notice that the nozzle tracking is severely affected by lighting, while resilient to image compression (mean error of approximately 2%).

TABLE 6.3: Effect of illumination and image compression on tracking accuracy (percentage error)

Test Case	maximum	mean	standard deviation
Lossless compression	4.78	2.50	1.40
Low compression	4.20	2.20	0.66
Medium compression	8.90	2.29	0.88
High compression	4.26	2.18	0.49
Good illumination	6.90	1.47	0.69
Mean Illumination	15.50	4.70	2.99
Bad illumination	23.90	8.53	6.40

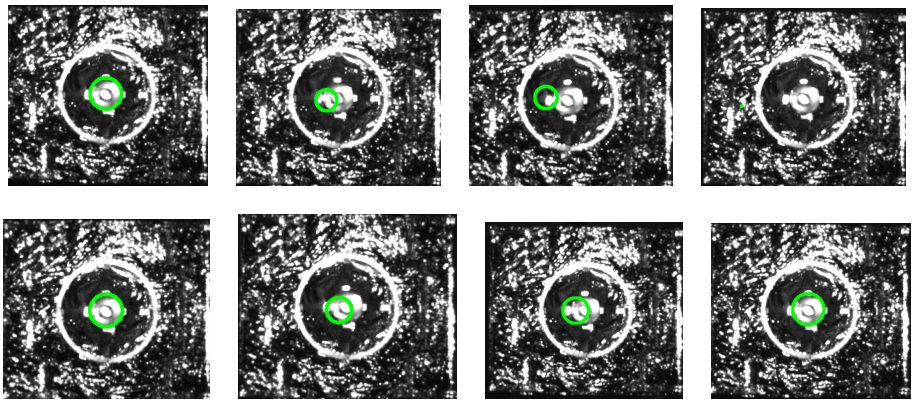


FIGURE 6.23: Integrating predictive simple motion model improves the performance of the tracking. The tracking without Kalman filter (first row) and the tracking integrated into Kalman filter (last row). The prediction prevents the loss of tracking.

6.5 Summary

On-orbit servicing may be performed on a client satellite that is controllable, even if some of its parts are malfunctioning. It is quite effective to exploit on-board sensors of the client for the approach phase of the on-orbit servicing task. In this regard, the attitude of the client satellite can be aligned with the servicer at the range where both attitude and position tracking are required. A vision-based system can be employed to estimate and track the position of the client relative to the servicer in 3 degrees of freedom (DOF), while attitude measurement is provided by the on-board attitude sensors. This is a great advantage for vision-based methods which may fail to recover 6 DOF pose in certain configurations of the client satellite (e.g symmetric and planar object).

In spite of a reduced DOF, position tracking is yet a challenging problem because of the severe space lighting and the reflective multilayer insulation of the satellite. This problem is not addressed in the literature despite a few attempts without considering the space lighting and reflective property of the surface of the satellite. In this section, we provide a novel hierarchical localization system from mid-range up to the contact point and prove the developed methodology.

The tracking of the client along a range of distances is handled by considering different image features. Silhouette contours with a dark background under poor lighting are difficult for tracking because of weak edges. Furthermore, the silhouette contours are invisible at close range in the field of the camera view. Therefore, we employed a

hierarchical model-based foreground and background segmentation at mid-range and nozzle tracking at close range.

At mid-range (20 m to 5 m), a simple model of the satellite is re-projected at a position hypothesis, while collecting the background and foreground histograms, and iteratively optimized using the simplex algorithm. The initial pose is determined by a brute-force search in a grid of boxes divided equally and refined by a simplex algorithm.

Once the servicer approaches the pre-determined close-range, the tracking switches automatically to edge-based nozzle tracking. The tracking then proceeds with a point to point matching, points sampled uniformly from the rep-projection of the 3D circle at a position hypothesis to the image edge along the edge normals. The translation parameters are estimated by iteratively optimizing using the Gauss-Newton algorithm. In case of mistracking, we employ a nozzle detector which globally searches and computes the nozzle position. For this purpose, the nozzle in a frontal view (circle) is detected by integrating RANSAC and absolute geometric error cost function, and refined by stereo constraints and the least squares optimization using the circle model. Furthermore, the Kalman filter is employed to handle a short-term poor lighting condition by introducing a non-linear motion model which reduces to linear constant velocity model for attitude-aligned satellites.

An extensive evaluation of the localization system was performed using by using one of the most realistic proximity operations simulator that reproduces the satellite surface properties and motion as well as the space lighting. The tracking at the mid-range is shown to be accurate and robust even with compressed images and poor lighting. A particular attention was given to the critical close-range approach where we thoroughly evaluated the effect of the sun direction, image compression and target illumination system on the performance of the tracking. The tracking performance depends on the direction of the sun and the satellite with respect to the camera axis. When the sun angle approaches to 90° with respect to the client satellite, a reliable tracking of the nozzle can be achieved with an additional source of light mounted on the servicer satellite because of the poor illumination and shadow, otherwise mistracking may occur which may be difficult to recover by detection. Therefore, a target illumination system is required to enhance the contrast of the image. The level of the target illumination is also critical and should be reduced as the sun angle approaches to 45°. On the other hand, the effect of

compression on the tracking is evaluated; the tracking can be performed without loosing much accuracy with a lossy compression of 12.5% and lossless compression of 50%.

Chapter 7

Model-based Pose Localization under Specular Reflection

In Chapter 6, we have considered a client satellite near end of its life and exploited on-board attitude measurement sensors to align the orientation to that of the servicer. Thus, a vision-based method is used only for estimation of the three translation parameters, while handling small attitude disturbance due to camera motion induced by the servicer. However, on-board attitude sensors are not always accessible. There are several satellites which are already fully non-cooperative, thus out of control and are freely tumbling or spinning in space. We rely on external sensors for the approach of the non-cooperative satellites. In this Chapter, a pose tracking of a tumbling satellite in six degrees of freedom (DOF) based on its CAD model and camera images is discussed. We adapt the state of the art edge-based tracking methods [4, 15], while improving with the search distance along the edge normal to match ambiguous edges [173] for long range tracking (Fig. 7.1 and Fig. 7.2). Multiple hypothesis tracking [74, 86] considers different hypotheses corresponding to potential edges to avoid problems due to ambiguities between edges. However, the multiple hypothesis solution may increase computational cost, and is limited to ambiguous edges at a certain pose, which does not address the problem for long distance tracking. Moreover, we speed up the model-based tracking by partitioning the CAD model of the satellite into its connected parts.

The pose tracking is presented in Section 7.1. The Section 7.1 is related to the Section 6.3.2 of Chapter 6, hence the details of some equations and related literature may

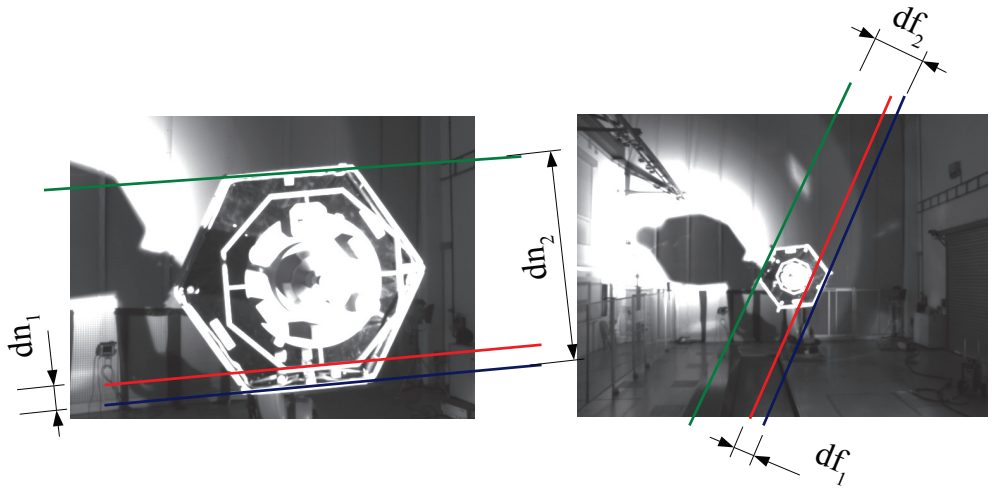


FIGURE 7.1: Ambiguous edges for matching along their normals. From far (right), the distance between two edges such as df_1 and df_2 is smaller than that of dn_1 and dn_2 from a close distance (left). A constant search length along the edge normals (lines perpendicular to the red, blue or green line) for both far and close distance creates ambiguity, i.e. the normals to the blue line may be also the normals to red line or green line particularly from far distance(right).

not be repeated here, therefore the reader may refer to earlier Chapter. Related to the pose tracking is pose detection, which is used to initialize the tracker at the beginning and in case of loss of tracking. While a satellite with several structures (edges) such as in Section 7.1 and 7.2 can be detected with the state of the art pose detection methods [174, 175], a satellite fully covered with the multilayer insulator (MLI) is a difficult and challenging problem. In Section 7.3, we address the problem of pose detection of a highly specular satellite by an appearance learning.

7.1 Pose Tracking in Six Degrees of Freedom

In this Chapter, we deal with a fully malfunctioning client satellite, without on-board attitude sensor and vision aid markers. We rely solely on vision-based methods to track the full 6 DOF pose (rotation and translation) of a non-cooperative client satellite. Moreover, we assume the CAD model of the client satellite is available, so that its absolute pose can be estimated.

The pose tracking methodology follows the same formulation as in Section 6.3.2, however in this case the number of parameters (six) to be estimated and visibility determination increase computational complexity. This is because we need to determine visible views and sample points in this views at a given pose hypothesis in real-time. In presence of a model of the object, pose tracking is achieved by projecting visible model edges at a pose hypothesis (last pose) onto the image plane, and aligning them with detected edges on the image, by searching along the edge normal and computing the minimizer of the cost function. The geometric model of the satellite is primarily used as a prior to predict the current motion at previous known pose for robust tracking. Real-time tracking can be achieved using an efficient implementation in a stand-alone computer, by partitioning a simplified CAD model of the satellite to its connected components. In fact, for space applications we rely on standard CPUs, simplifying the complex model provided by the original design files, in order to remove tiny surface details that are invisible at the desired range of distances, so that we could sample visible model lines with real-time efficiency.

The pose estimation and tracking approach is generally similar to Section 6.3.2 and minimizes the re-projection error

$$\hat{\mu} = \arg\min_{\mu} \sum_{i}^{n_p} ||\mathbf{s}_i - \mathbf{y}_i(\mu)||^2 \tag{7.1}$$

where $\mathbf{y}$ is the projection of visible 3D sample points onto a camera image plane at pose parameters μ, $\mathbf{s}$ is 2D image point of the nearest edge corresponding to the projected 2D point of the model, and n_p is the number of valid sample points respectively. The Jacobians can be easily computed from the re-projection error, and the pose parameters μ are determined using Gauss-Newton optimization, iteratively refining the pose estimate until it converges. In this case, we estimate the 6D motion parameters while the client satellite is in arbitrary motion. Arbitrary motion in turn calls for visibility determination during the tracking. Thus, before projecting model edges onto the image plane, it is required to determine which part of the CAD model is visible at the last pose.

7.1.1 Visibility Determination for Pose Estimation

A rotating object changes views from the camera. If we project the 3D model of the object onto the image plane at a pose hypothesis, the corresponding 2D image may not match at all with observed camera image. In tracking, the pose hypothesis is mostly a pose estimate from the last frame. Therefore, we first determine which faces of the object are visible at last pose, and it is likely that most of those faces are visible at the current pose.

Unlike attitude-aligned client satellites, the computation of the visible geometry from a view point is crucial for model-based tracking of a freely tumbling satellite. This computation is normally expensive and prohibits real-time tracking. GPU based methods are widely employed to accelerate the computation [86]. Alternative to GPU based methods is the parallelization with multi-core computers. However, the standard in space applications is mostly restricted to single thread programs running on CPU due to safety. In this thesis we limit ourselves to the current space standard. For this purpose, thanks to a part-based model representation and a model reduction, the visibility of a certain face of the model can be quickly determined by computing the Z-buffer using CPU and without parallelization. The Z-buffer is a simple hidden surface removal method, which is used to determine visible faces during tracking. A Part-based model representation is a description of an object model with one or several parts of the object which are connected, i.e. instead of the single CAD model of the object we use several parts (structural elements composing the model). This representation speeds up the determination of the visible views, because the Z-buffer rendering is restricted only to the relevant part of the model that is closest to the last view or pose.

7.1.2 Adaptive Search Distance

As described in Section 6.3.2 of Chapter 6, the model-based tracking relies on searching the correspondences along the perpendicular component of the motion because of the aperture problem [4, 15, 59]. The appropriate search length along the edge normal requires to be selected, and fixed once it is selected. This is usually sufficient for small range tracking, as in industrial application. However, a constant search length along the edge normal may not be robust for tracking in long range.

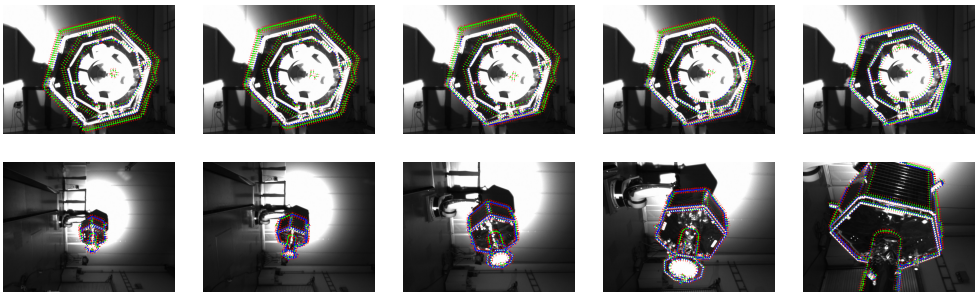


FIGURE 7.2: Pose estimation process. Top row: Iterative (here shown only six iterations) optimization along edge normals. Bottom row: Tracking at far (7 m) to near distance (2.5 m). The distance between two edges in the image varies according to the range (see first image and the last image of the same row), therefore variable search distance along the edge normal is necessary.

Therefore, in order to alleviate this problem we employ an adaptive search length based on the range of the target satellite to the camera and size of the object. The concept of variable search distance is also mentioned in [13], although it is not described how it could be adaptive. Here we assume that the camera lens system is properly selected according to the distance of the object from the camera. Let the maximum length of the object be L_o, the distance of the object from the camera d and the focal length of the camera f, then the maximum size of the object in the the image is given by

$$l = \frac{fL_o}{d}, \tag{7.2}$$

which can be easily derived from similar triangles of the geometry of image formation. The search distance along the edge normal is proportional to the object size l in the image plane. The proportionality constant may be determined from the slope of the curve in the Equation (7.2). The slope of the function $l(d)$ describes how fast the curve given by graphing image size versus distance falls. This fall of the curve can be an indicative of variation of the search length in the image against range of the object from the camera (Fig. 7.3).

7.1.3 Pose Prediction with Kalman Filter

The pose tracking at current frame depends on the previous pose; it is assumed that the edges from the previous frame do not move significantly. This assumption often holds true for smooth motion and at low speed of a satellite. However, the pose tracking is prone to failure at large inter-frame motion of the satellite. Moreover, a single frame

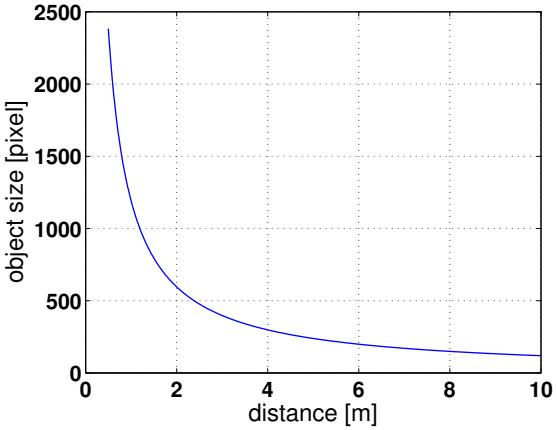


FIGURE 7.3: The relationship between the size of an object in image plane and the object distance to a camera. The maximum size of the object here is 2 m and focal length 6 mm.

drop-out due to image saturation could not be handled. The better approach is to predict the pose at the next frame using motion model of the satellite. We assume a constant velocity motion between two camera frames and formulate the pose prediction in terms of inter-frame motion parameters in the framework of Kalman filter.

The Kalman filter provides:

- a frame-to-frame prediction capability for pose tracking of a satellite,
- reduction of noise by temporal integration of the pose estimate.

The inter-frame motion $\mathbf{v}$ at time interval Δt and the current known pose E_k is used to predict the pose $E_{k+1,k}$ at the next frame as $E_{k+1,k} = exp(\mathbf{v}\Delta t)E_k$. The states are represented as an inter-frame velocity vector $\mathbf{v}$ and the associated transformation matrix E_k. This is an equivalent to the classical EKF, where the pose parameters are represented by $Log_{SO(3)}(E)$. The former representation effectively linearizes the filter around the current pose estimate and simplifies the calculation [176].

The filter estimate $\mathbf{x}$ of the true state is given by

$$\mathbf{x_k} = \{exp(\mathbf{v}_k\Delta t)E_k, \mathbf{v}_k\} \tag{7.3}$$

and the state estimate E_k is related to the true state $\bar{E}_k$ as

$$\begin{aligned} \bar{E}_k &= exp(\boldsymbol{\rho}_{pose}) E_k \\ \bar{\mathbf{v}}_k &= \mathbf{v}_k + \boldsymbol{\rho}_v \end{aligned} \tag{7.4}$$

where the state error vectors $\boldsymbol{\rho}_{pose}$ and $\boldsymbol{\rho}_v$ at discrete time k can be modeled as multivariate normally distributed Gaussian noise of zero mean and state error covariance $\boldsymbol{\Sigma}_k$. The process noise, which is deviation from the constant velocity model (acceleration) has a state noise covariance matrix $\mathbf{Q}_k = diag(\boldsymbol{\Sigma}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\omega})$, with $\boldsymbol{\Sigma}_{\mathbf{v}}$ and $\boldsymbol{\Sigma}_{\omega}$ are respectively diagonal covariance matrices associated to the standard deviations of the translation and rotation parameters.

The pose prediction is based on the constant velocity model over a time interval Δt, with the corresponding transition matrix F. The state covariance is determined by

$$\boldsymbol{\Sigma}_{k+1,k} = F \boldsymbol{\Sigma}_k F^T + \mathbf{Q}. \tag{7.5}$$

The measurements are provided in the form of pose estimates of the transformation matrix E^m with measurement covariance R. The residual covariance (innovation) S which describes the motion from a priori state to the measurement pose is

$$S = E^m_{k+1} E^{-1}_{k+1,k} \tag{7.6}$$

where $E_{k+1,k}$ is the predicted pose (a priori) and E^m_{k+1} is the current pose measurement in matrix form.

The Kalman gain is given by

$$K_k = \boldsymbol{\Sigma}_{k+1,k} H^T (H + \boldsymbol{\Sigma}_{k+1,k} H^T + R)^{-1} \tag{7.7}$$

where the observation model $H = [I_6 \quad 0]$. Finally, similar to [176] the a posterior estimate E_{k+1} is determined using a priori pose, and weighting the innovation by the kalman gain

$$E_{k+1} = exp(K log(S)) E_{k+1,k} \tag{7.8}$$

and the covariance is

$$\Sigma_{k+1} = (I_6 - K_k H)\Sigma_{k+1,k}. \tag{7.9}$$

The Equation (7.8) gives the smoothed pose estimate at the current frame based on the state prediction (a priori) from the last pose and the current measurement. The uncertainty of this pose estimate is given by Equation (7.9), which could provide a measure of reliability of the pose tracking at each camera frame.

7.2 Evaluation

We evaluate the 6 DOF pose tracking against the ground truth generated both with a robotic hardware-in-the-loop and a computer graphics-based simulator described in Chapter 4. The hardware-in-the-loop simulator at the European Proximity Operations Simulator (EPOS) facility of DLR is used for evaluation of rendezvous and proximity operations. Here we use the EPOS facility to validate the pose tracking method described in Section 7.1 with real image sequences. Moreover, we evaluate the algorithm with photo-realistic images generated through a ray tracing with more accurate sun simulator and ground truth. Notice that the hardware-based experiment reproduces the multilayer insulation (MLI) of the satellite accurately, while the computer graphics simulates the sunlight more precisely than high power floodlight. Therefore, evaluating the tracking algorithm with both methods could complement the shortfall. We focus more on the hardware-based test platform because the accurate modeling of the MLI highly influences the tracking.

7.2.1 Results on a Hardware-in-the-Loop Simulator

The motion of a tumbling target satellite (Fig. 7.4) is simulated with a 6 DOF robot motion. The goal of this evaluation is to verify the model-based tracking described above under various motion and illumination. Two sun directions are considered in this experiment (0° and 45°) for all motion profiles. The first type of motion considered is in-plane roto-translation; the client satellite undergoes a planar motion (rotation and translation mainly about/along the optical axis). The second type of motion is out-of-plane roto-translation at a nutation angle of 45° about the yaw-axis (vertical axis),

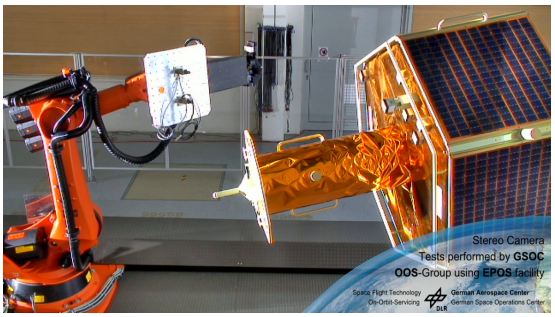


FIGURE 7.4: A mock-up of a target satellite (right) mounted on a six degrees of freedom (DOF) robot and stereo camera mounted on another six DOF robot-servicer (left). The target satellite undergoes a tumbling motion while the servicer approaches it.

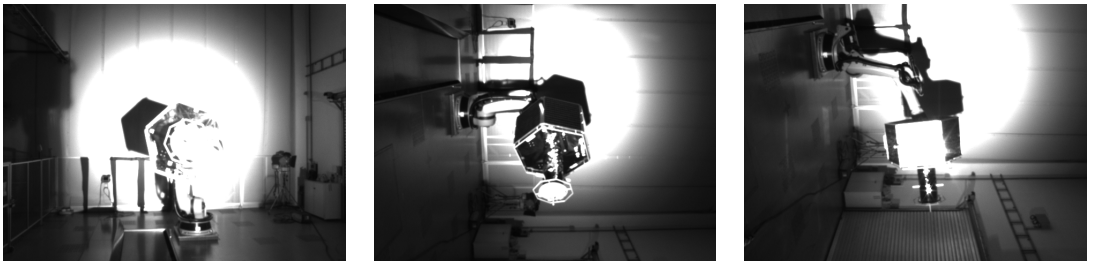


FIGURE 7.5: Images taken at the same distance to show the motion types used for evaluation. The in-plan motion frame (left), out of plane motion at a nutation angle of 45° (middle) about yaw axis and out of plane motion at a nutation angle of 90° and 10° (right) about yaw- and roll-axis respectively.

and the last motion considered is out-of-plane motion at a nutation angle of 90° about yaw-axis and 10° about the roll-axis (see Fig. 7.5). These are exemplary trajectories reported here. Indeed, more trajectories have been tested by varying the angular and linear speed slightly (maximum, 4 deg/s and 50 mm/s respectively). However, in our experiment the influence of the satellite speed on the tracking was insignificant because of the assumed small maximum speed.

A stereo camera of image resolution 640×480 and a focal length of 6 mm is used to record image sequences. The baseline between the two stereo camera is 28 cm. The average processing time of a monocular image with Intel Xeon (R) 2.8 GHz CPU is 100 ms. The performance analysis is based on the ground truth trajectories obtained from the measurements of the two six DOF robots at EPOS facility, German Aerospace Center (DLR). Furthermore, we estimate the transformation from the Tool Center Point (TCP) of the robot to the camera frame through a hand-eye calibration procedure. Thus, we obtain the ground truth poses through robot measurement and hand-eye calibration.

Regarding the experimental results, we evaluate the effect of the sun direction for the

same motion profiles (in-plane, out-of-plane and a large nutation out-of-plane roto-translations) at the same illumination condition. Both rotation and translation errors are higher for the in-plane motion at the sun direction of 0° (Fig. 7.6), because the client satellite is frontal to the sun light and several edges in the image are over-illuminated resulting in less features for matching (see images in Fig. 7.7). In contrast, the tracking accuracy improves when the sun rays strike the satellite in frontal view at an angle (45°). Notice that the error along the optical axis is exceptionally higher and comparable for both cases in the first few meters. One of the reasons may be the fact that an accurate estimation of the translation of an object in frontal view along the optical axis is difficult in both illumination directions.

In contrast to in-plane motion, the translation errors are higher for the out-of-plane motion at the sun direction of 45° (Fig. 7.8). This is because the client satellite changes view, resulting in shadow and under-illuminated edge features (see images in Fig. 7.9). In contrast, the tracking accuracy improves at an angle of 0° due to direct light in the direction of camera line of sight. This effect appears to be true for a tumbling satellite which changes views and create self-shadow, however slightly exaggerated because of the non-parallel ray of the sun simulator.

On the other hand, the rotation and translation errors are slightly higher for the out-of-plane motion with large nutation angle (90°) at the sun direction of 45° (Fig. 7.10). However, in this case due to large the nutation angle (Fig. 7.11) the satellite drifts from the main rotation axis, and some views become parallel to the sun direction of 45°. Thus, the effect of the two sun directions in some satellite poses is equal, resulting in a comparable error.

The above discussion and comparisons are based on the evaluation of the pose tracking, which relies on the adaptive search length for the long range (in our experiment up to 7 m) and Kalman filtering for large inter-frame motion. At close-range (5 m to contact point), the variable search distance does not influence the pose tracking.

7.2.2 Results on Photo-realistic Rendered Images

The pose tracking is also evaluated with photo-realistic synthetic images (Fig. 7.12) rendered from the 3D model of a full scale satellite. Despite the differences in the

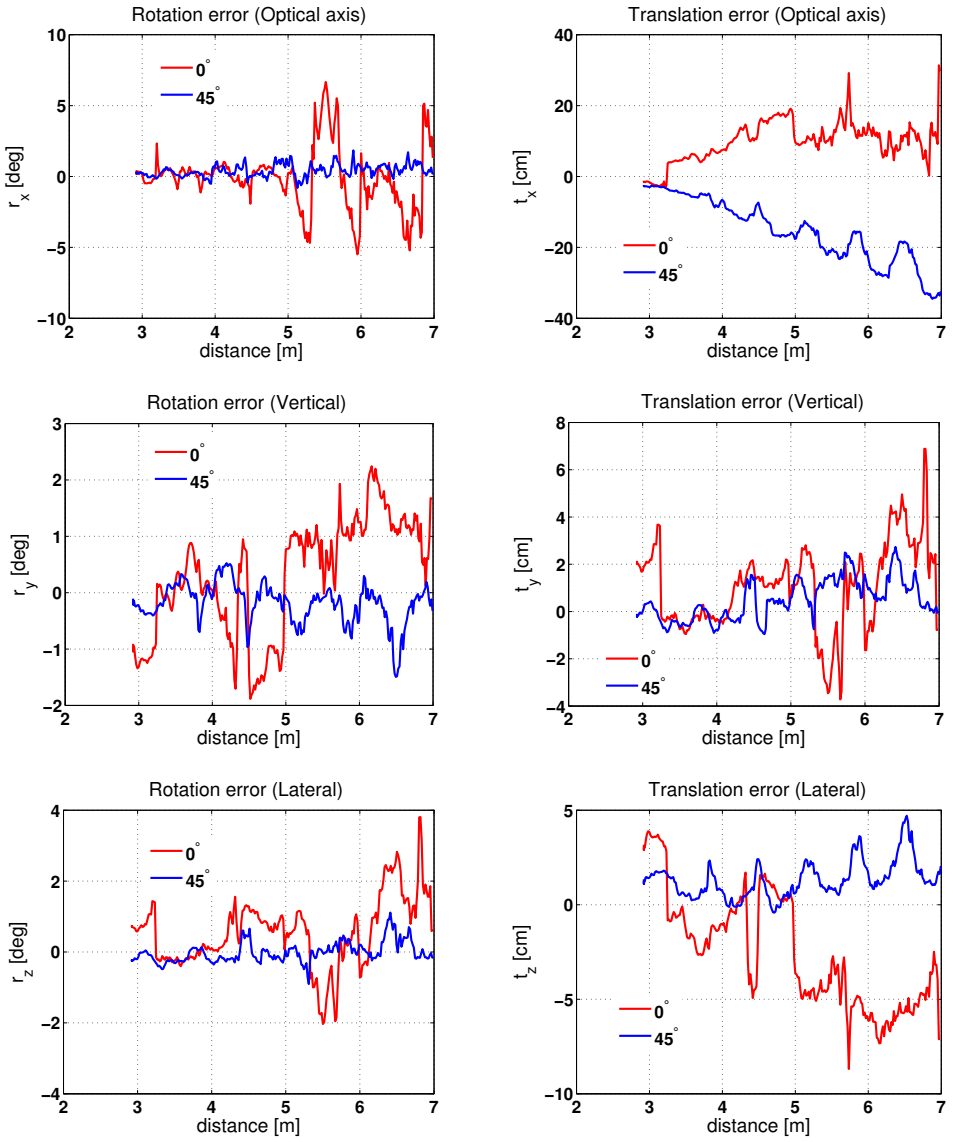


FIGURE 7.6: Effect of the sun direction on pose tracking for in-plane roto-translation motion. The sun simulator was placed at angles of approximately 0° and 45° to the camera optical axis.

structure of the satellite, the same motion trajectory is used to evaluate the ground truth error. In this case the ground truth and the sun simulator are more accurate than hardware based simulation. However, the simulation of the multilayer insulation was difficult and less accurate. Here, we present the results of the pose tracking for two motion trajectories. Both motions are out-of-plane roto-translation but at different nutation angles. The angle of the sun in both cases is 90° with respect to client satellite local vertical local horizontal frame (LVLH).

Fig. 7.13 shows the pose tracking error evaluated with the synthetic photo-realistic

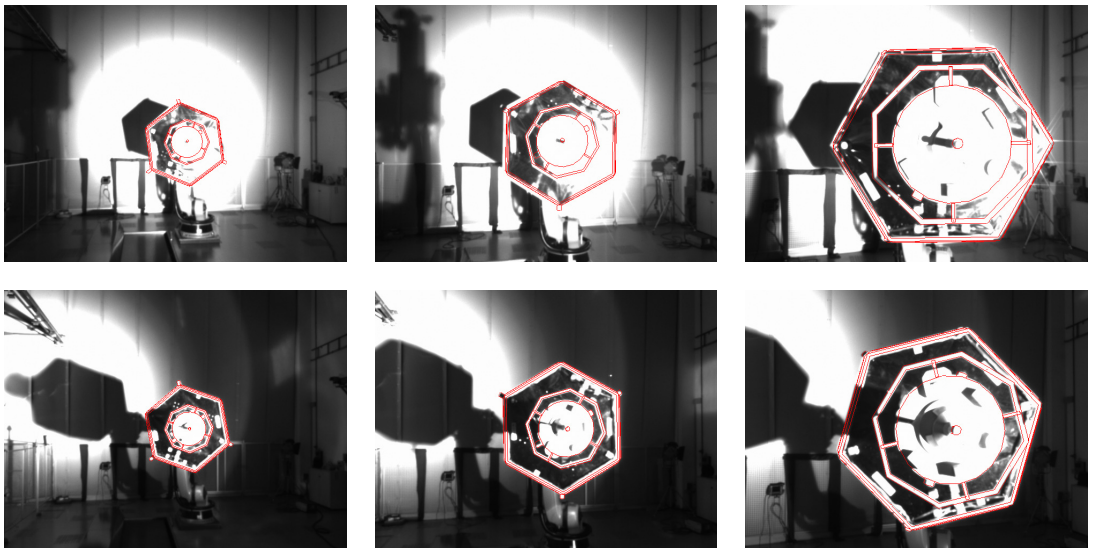


FIGURE 7.7: Image sequences of in-plane motion for the sun direction 0° (first row) and 45° (second row).

images (Fig. 7.14) for the out-of-plane roto-translation motion with a nutation angle of 45°.

The computer graphics and hardware based evaluations here may not be compared as they are not at the same condition in terms of the intensity of light and satellite geometry. However, the error distribution of the tracking with rendered images appears similar to that of the hardware based evaluation (Fig. 7.8 and 7.13). This similarity may have arisen because of the same motion trajectory and the quality of rendered images.

Similarly, the tracking error of the out-of-plane roto-translation with a larger nutation angle of 90° is shown in Fig. 7.15. One may question which evaluation best describes the pose tracking error. We assert that a hardware-based evaluation is more reliable than the photo-realistic rendering because the reflective MLI influences the performance of the image based tracking significantly. In this regard, the satellite mock-up reproduces the MLI more accurately than computer graphics based modeling. However, the photo-realistic rendering is useful to analyze the effect of the sun direction without any constraint.

7.2.3 Summary

A fully malfunctioning client satellite may undergo spinning and tumbling motion. For an approach and a manipulation of a tumbling satellite, we rely on pose estimation and

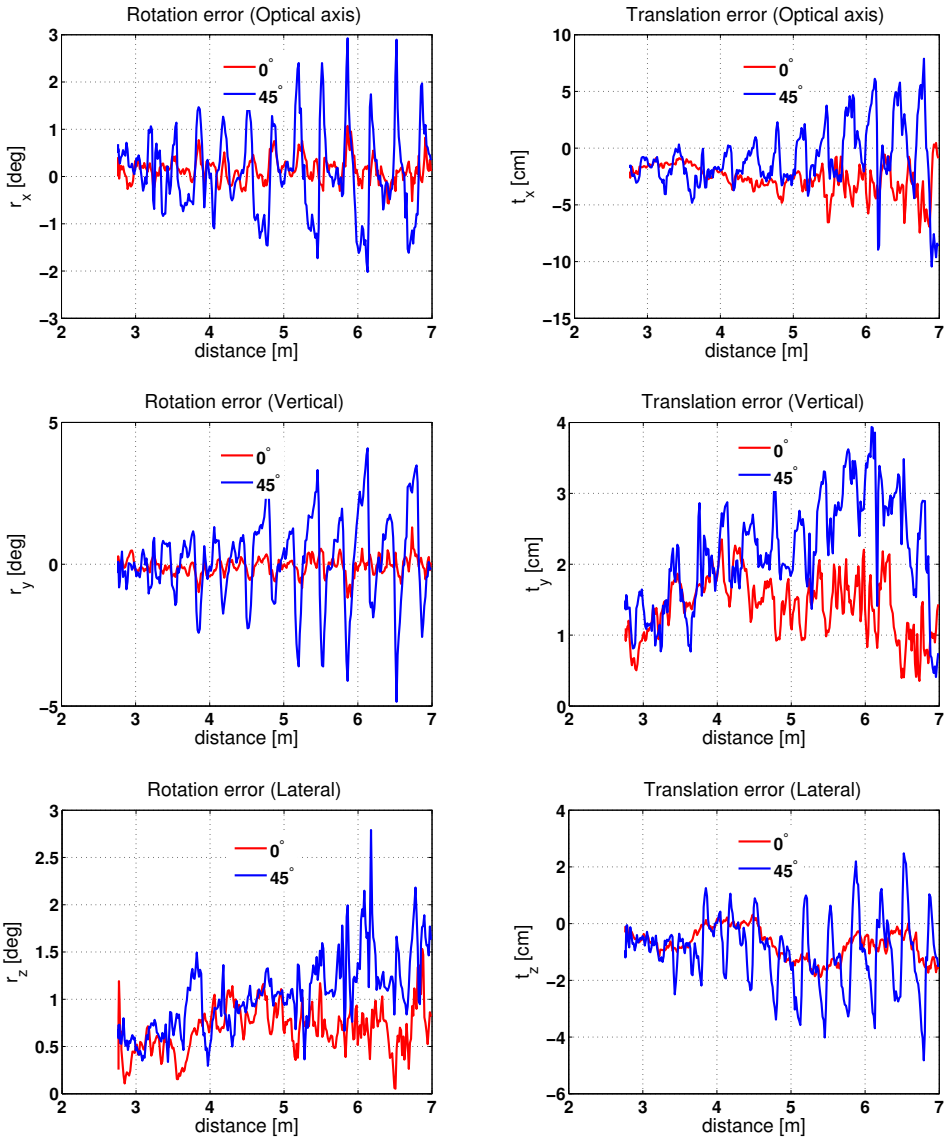


FIGURE 7.8: Effect of the sun direction on pose tracking for out-of-plane motion. The sun simulator was placed at angle of approximately 0° and 45°.

tracking in six DOF. Mostly we have the knowledge of the CAD model of the malfunctioning satellite during design. The CAD model of the satellite is a great advantage for the tracking and pose estimation. The pose estimation is achieved by aligning the visible model and image edges along edge normal at a given pose hypothesis, provided by the last pose. However, matching edges by 1D search along the edge normal from far distance creates ambiguity for closely located image edges. To alleviate this problem, we implemented an adaptive search distance along the edge normal that depends on the distance and object size.

The pose tracking in a six DOF are evaluated based on the realistic proximity operations

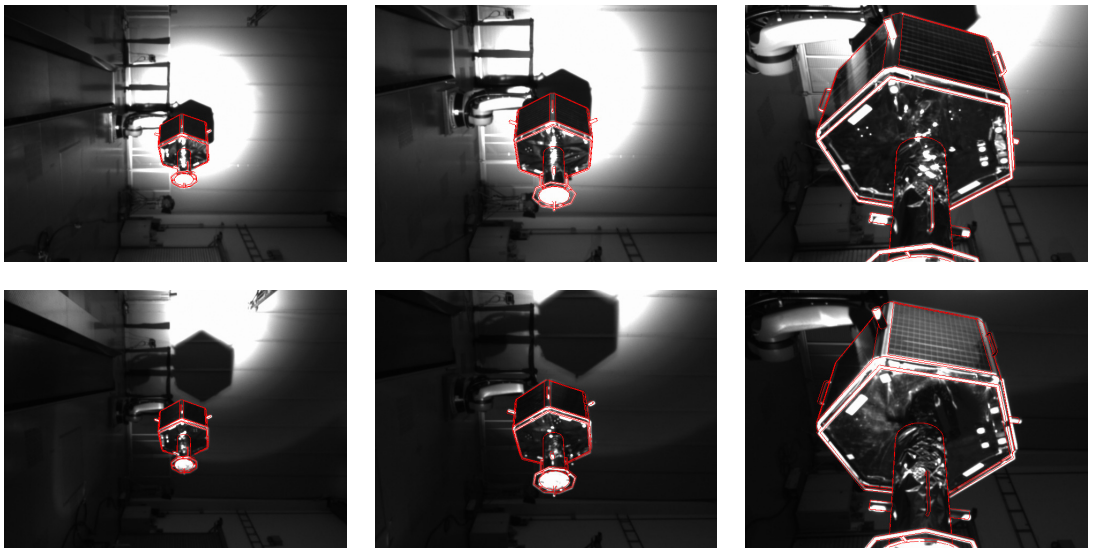


FIGURE 7.9: Image sequences of out-of-plane motion at the sun direction 0° (first row) and 45° (second row).

hardware facility. The angle between the sun and the camera line of sight (the sun direction) with respect to the client satellite significantly influences the tracking accuracy. The experimental results indicate that the performance of the tracking depends on the direction of the sun as well as the type of motion (rotation). The tracking error increases for in-plane motion (satellite in frontal view) as the sun direction approaches to 0° due to over-illumination. In contrast, the tracking accuracy decreases for out-of-plane motion (satellite nutates from main axis) as the sun direction approaches 45° because of the shadow and poor illumination. On the other hand, as more nutation is introduced to two of its axes for the out-of-plane motion, the pose tracking accuracy barely decreases at the sun direction of 45° because of a large change of view.

On the other hand, a photo-realistic rendering, which simulates the sun spectrum more accurately based on computer graphics, was used to evaluate the pose tracking. Despite imprecise simulation of the MLI, a photo-realistic image can be used to evaluate the tracking methods at more accurate sun simulation.

7.3 Pose Detection of a Satellite

Detection of a satellite in our context refers to determining a region of interest (satellite) in an image. Here we assume the image contains a single satellite, and we estimate position and orientation (pose) of the satellite in 3D space based on images and CAD

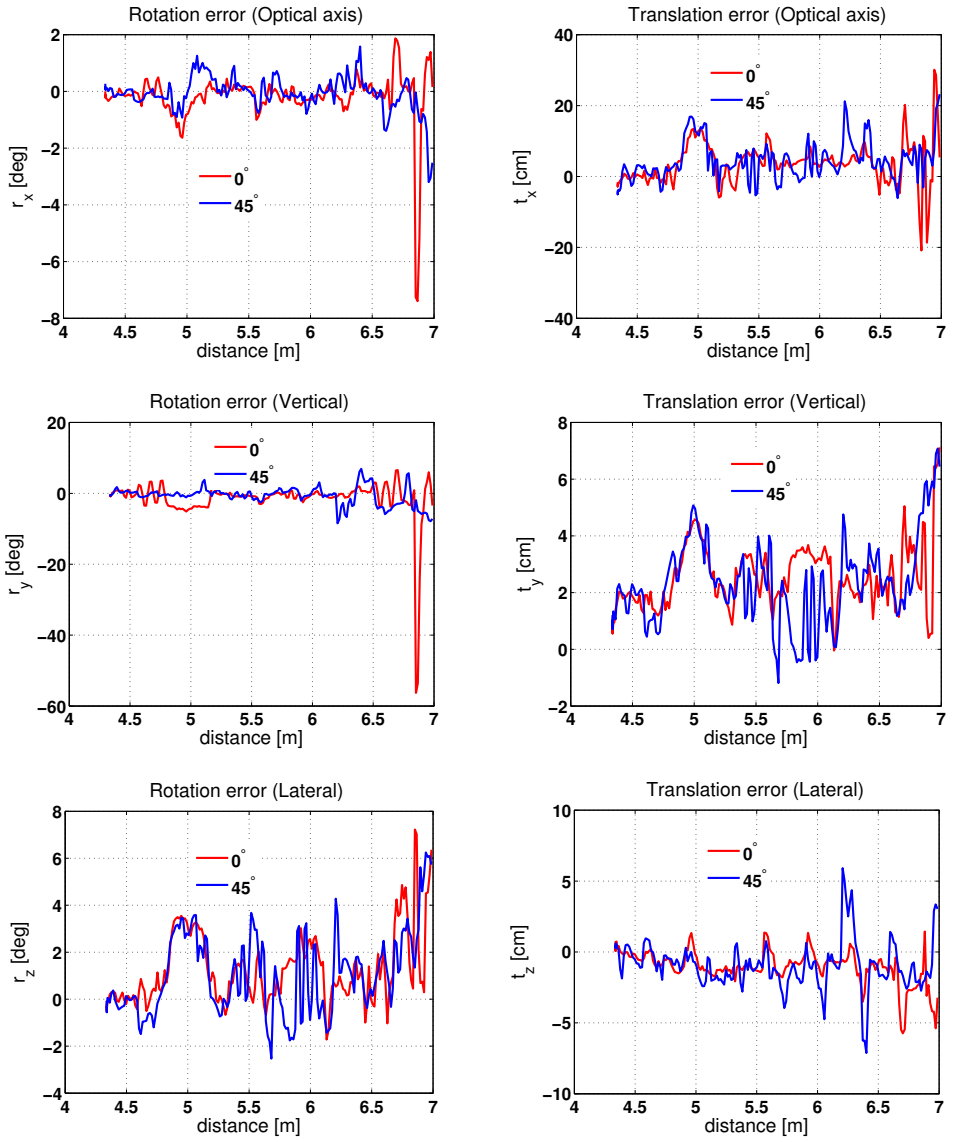


FIGURE 7.10: Effect of the sun direction on pose tracking for out-of-plane motion and a nutation angle of 90°. The sun simulator was placed at angle of an approximately 0° and 45°.

model. Pose detection is generally used for the initialization of any tracking method, which is based on a local optimization. A global detection of an object poses difficulties because of the large search space in six DOF. This global search is even worse in the case of a satellite pose estimation due to the harsh space lighting. Here we assume further that the specific location (orbit) of the desired satellite in space is pre-determined from ground or space based observation, i.e. no more object recognition is required .

There exist a number of object detection methods. View-based approaches [174, 175, 177, 178] compare a search image with the 2D views of the object precomputed from

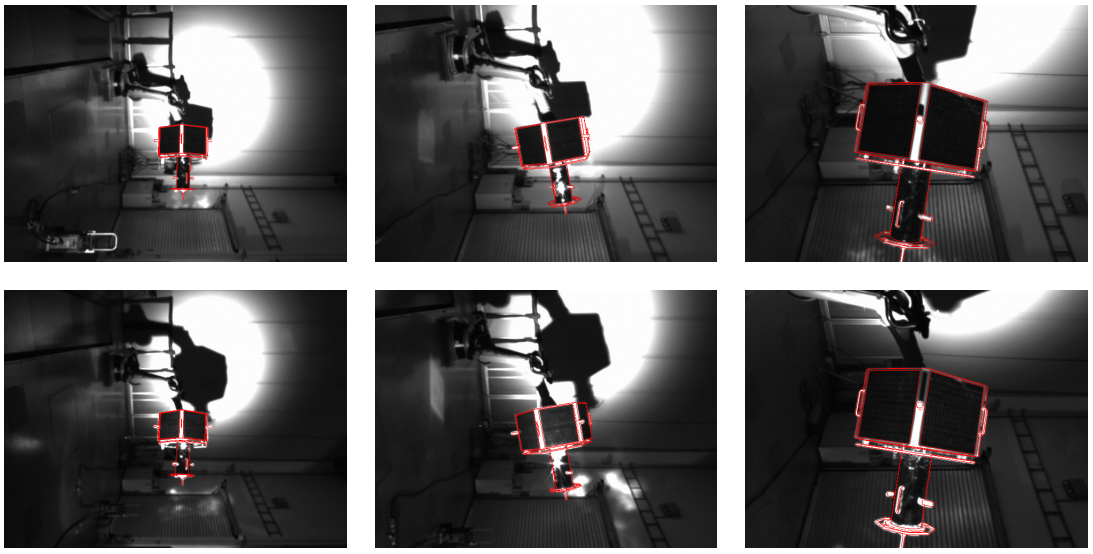


FIGURE 7.11: Image sequences of out-of-plane motion with a nutation angle of 90° at the sun direction 0° (first row) and 45° (second row)

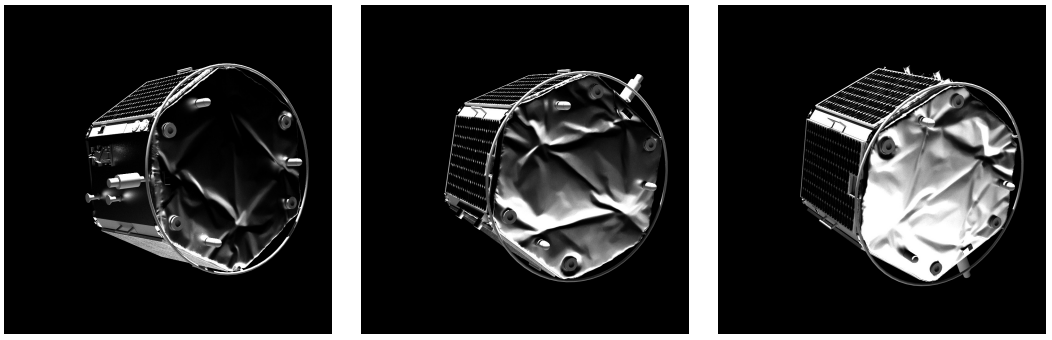


FIGURE 7.12: Mock-up of the satellite in various views used to evaluate pose tracking with a computer graphics based sun simulator.

its 3D model by clustering views. Related to these methods, [179] uses no training templates instead adapts the template through on-line learning. The view-based methods represent the views with an aspect graph either through a hierarchical clustering [175, 177, 178] or pyramid search of images [174]. [175] employed a foreground and background segmentation of the image as a pre-processing step, assuming static background and a few sequence of images. The experimental results demonstrated in [174, 175] appears relevant to a 3D detection of a space object, however such methods require prominent edges of the object to succeed in accurate pose detection.

On the other hand, in order to tackle the problem of a large search space, feature-based approaches [180–183] exploit image features such as corners, lines, intersection of lines and complex features from grouping of various image features. The corresponding 3D

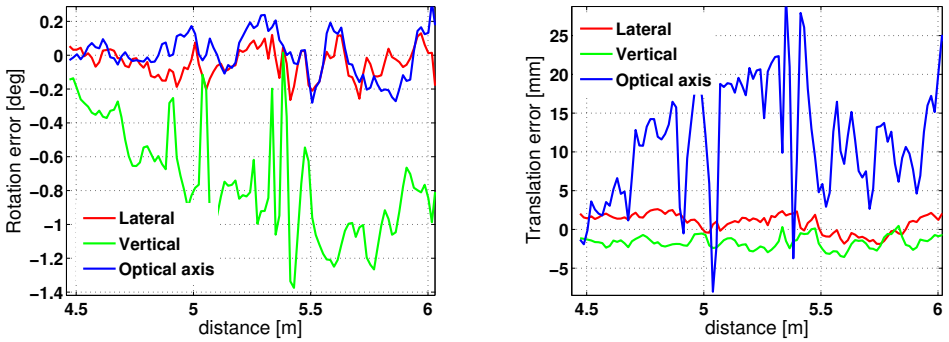


FIGURE 7.13: Pose tracking error with photo-realistic rendered images. The trajectory is the same as that of Fig. 7.8, i.e. roto-translation with a nutation angle of 45°. Notice that the client satellite in this case is different, and shown in Fig. 7.12.

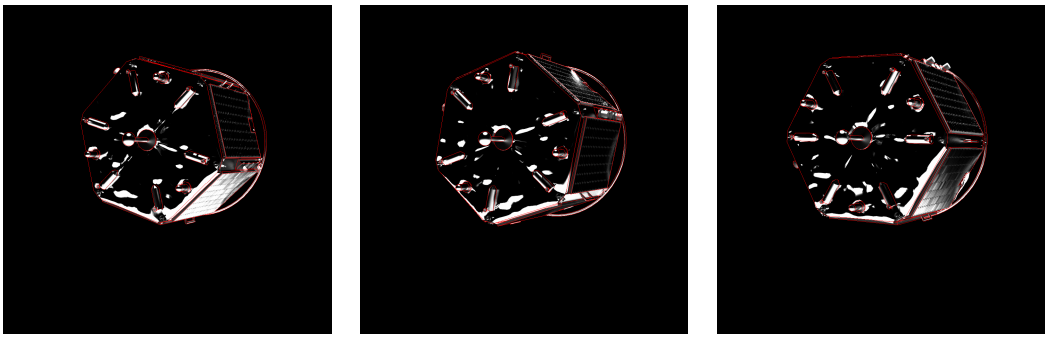


FIGURE 7.14: Photo-realistic image sequences of roto-translation with a nutation angle of 45°.

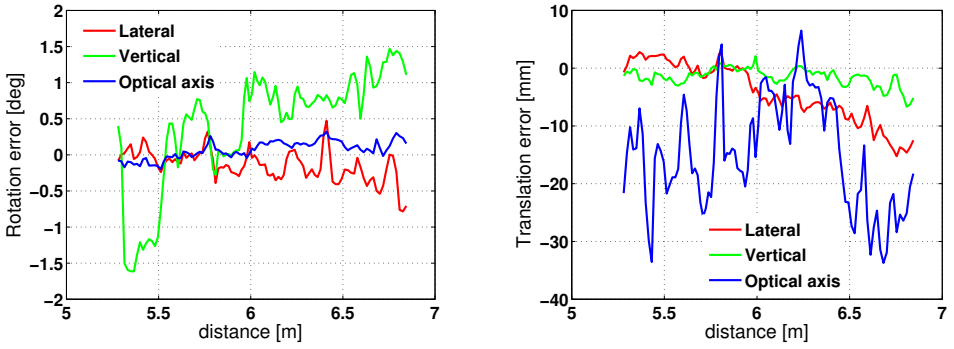


FIGURE 7.15: Pose tracking error with the photo-realistic rendered images. The motion trajectory is a roto-translation with a nutation angle of 90°, as in Fig. 7.10. The client satellite is shown in Fig. 7.12.

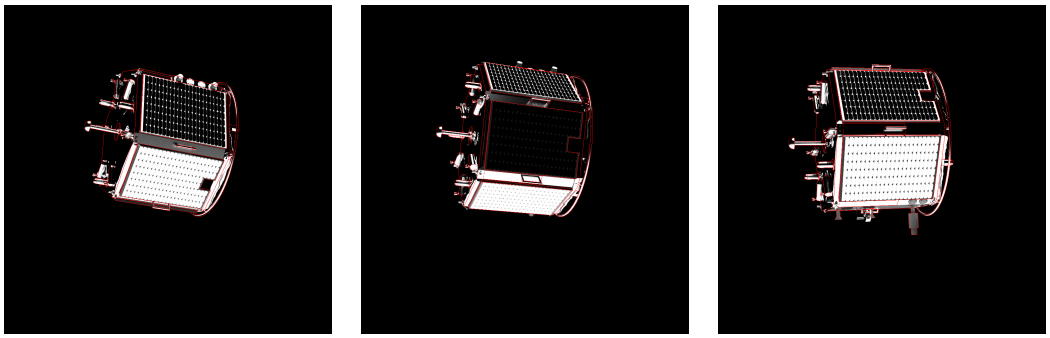


FIGURE 7.16: Photo-realistic image sequences of roto-translation with a nutation angle of 90°.

features are assumed to be on the 3D model of the object in consideration. Thus, the extracted image features (edges, corners and lines) are matched to the 3D features, and the 3D pose is computed. The main drawback of such method lies on dealing with the possible large search space related to establishing the correspondence between the image and object features. Moreover, background clutter in the image makes the feature extraction difficult.

Descriptor-based methods such as [184] learn local or semi-local features extracted from training images. The advantage of the descriptor-based approaches is that their run-time is independent of the size of the geometric search space. They show a good performance in recognition of textured objects. [184] uses at least one training image, while many others are generated through computer-graphics techniques such as texture mapping and rendering of the 3D model. However, the texture mapping from a given image and 3D rendering of specular object under a direct sunlight is inaccurate.

We provide a novel appearance-based pose detection method, which learns both lighting and view points, creating a large database of feature points to cluster and learn in a vocabulary tree based on [185]. The observed image is then queried in real-time and the corresponding pose is computed based on descriptor matching. The training images are obtained from a camera system mounted on six degrees of freedom robot. The operational workspace of the robot is limited, which poses difficulty to reach all the necessary view points. Therefore, we employ a 3D image warping to complement those missing views because of the robot workspace constraint during acquisition of training data.

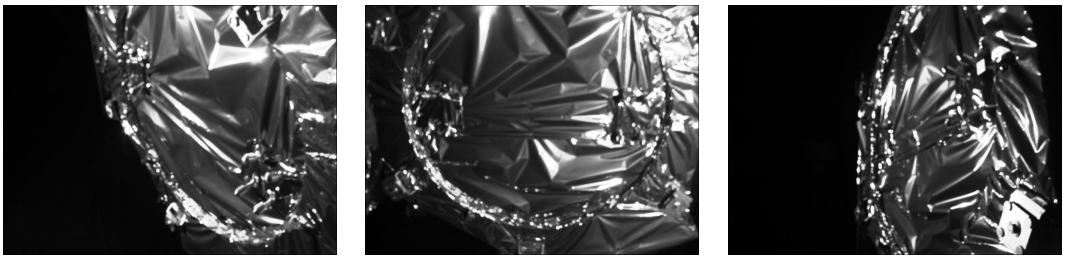


FIGURE 7.17: Images of the TerraSAR-X mock-up (rear part). The satellite surface is fully covered with a multilayer insulation (MLI) and consists of weak edges, difficult for pose detection.

7.3.1 Pose Detection by Appearance Learning

The appearance of a satellite changes significantly with the change of the direction of the sun because of the reflective property of the surface. In the absence of strong edges, detection of the satellite with edge templates under space lighting is difficult and hard to achieve the desired accuracy for an initialization of a local tracking. The detection of the satellite with distinctive edges can be achieved as demonstrated in [175]. However, when the most surface of the satellite is covered with a multilayer insulation (e.g. TerraSAR-X in Fig. 7.17), it is hardly possible to detect the pose using edge based methods because of several clutters created by the MLI and severe illumination.

In contrast, the wrinkles of the MLI and satellite structures such as blobs and junctions provide useful cues for the detection of the satellite. Indeed, we exploited such features for tracking in the absence of the CAD model of the satellite in Chapter 5. The main disadvantage of such point features is a drift in tracking over a long sequence; this is a general problem in a model-free feature tracking and even worse with a strong change of lighting. We are here concerned with the global detection of the satellite pose, which relies on a feature description of a region around each point feature. The feature drift unlike the model-free tracking is not an issue because the pose detection is not used in a long sequence, instead only to (re)-initialize in case of loss of tracking. Moreover, we assume the CAD model of the satellite exists, therefore a drift-free model-based detection as well as tracking is ensured.

Therefore, we employ a feature based pose detection of the satellite by computing a descriptor vector for each region of interest. We adapt an appearance learning, in which we learn appearances of the satellite under various view points and lighting directions. To efficiently retrieve the observed image in a large number of views stored in a database,

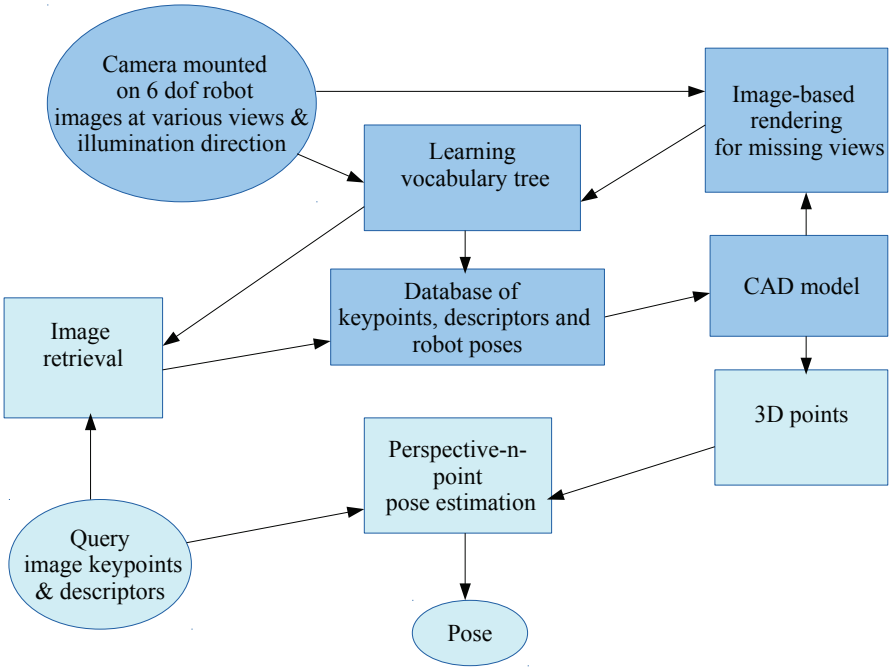


FIGURE 7.18: The overview of the pose detection method. The blocks in the diagram represent offline processing (in blue) and the online processing (pale green).

the appearance (feature) learning builds up on the popular technique of the vocabulary tree [185]. The representative keypoints from several views and different direction of the sun can be incorporated into the learning scheme. In the context of satellite pose detection, the strength of such approach is its robustness against occlusion, background clutter and illumination.

The work-flow (see Fig. 7.18) for the pose detection in the presence of the CAD model of the satellite is as follows:

- sampling view points at different sun direction,
- feature detection and description,
- descriptor quantization in a vocabulary tree,
- depth map computation from the CAD model,
- image retrieval and pose estimation by 3D-2D registration.

Photo-realistic training images can be obtained in two different ways depending on the existing information about the client satellite. In either case, we assume there exists a CAD model and a number of image archives of the satellite are readily obtainable. The CAD model of the satellite consists of geometric parts in a mesh form, and no material properties such as multilayer insulation (MLI). In the perspective of image processing, the MLI is highly determinant of the performance of the tracking, therefore has to be modeled accurately. Reproducing the specular property of the MLI is very difficult and by itself a research area in computer graphics. One way to obtain training images is the photo-realistic rendering based on a ray-tracing. The distant parallel sun light can be simulated with the computer graphics accurately, and the MLI can be modeled by texture mapping of the image of the satellite and its 3D model. However, the texture mapping of a reflective specular surface may not accurately reproduce the original material and optical property of the satellite surface, hence represents less reliable training data.

Therefore, we reproduce the MLI more accurately with a hardware (a satellite mock-up) for training images. The MLI of the satellite can be reproduced by taking account of the geometry and material property. The advantage of the satellite mock-up is that the true MLI can be wrapped around the surface of the satellite. We take the advantage of this method to get training images with the mock-up of the TerraSAR-X satellite. View points are first generated by placing the object in a center of virtual sphere and moving the camera around the surface of the sphere at an interval of 8° in latitude and longitude. The trajectories generated in this manner, but in a hemisphere were implemented using a real stereo camera mounted on a six DOF robot. The robot moves in pre-planned trajectories, repeating those views at different sun directions, where the actual sunlight was simulated with a high power flood light.

The disadvantage of this method is that the workspace of the robot is constrained, and it was not possible to reach the desired view points. In order to address this problem, we employ an image-based rendering technique based on the 3D warping. A virtual camera image at a desired view point can be synthesized from at least one reference image, by back-projecting each pixel of the reference views using the depth map and re-projecting to the new view, as shown in Fig. 7.19.

An image rendered by the 3D warping has problems of holes or gaps due to round off error resulting from the decimal to integer conversion of pixels, magnification when the

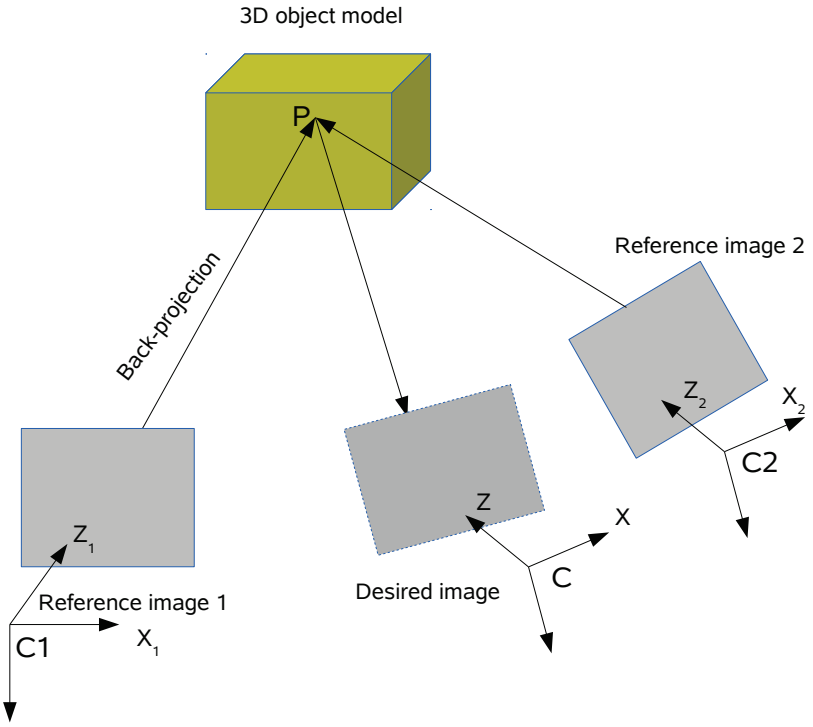


FIGURE 7.19: Image-based rendering using the 3D warping. The reference image pixels are back-projected to the 3D space using depth map computed from the 3D model of the object. The 3D points P are then re-projected onto the desired view to synthesize the desired image.

camera moves closer and depth discontinuity causing dis-occlusion [186]. Post-processing the newly synthesized image has been proposed to handle the problems of gaps or holes due to sub-pixel location of the desired image. However the post-processing blurs the image, which is an undesirable property for feature-based matching. We employ a simple but effective morphological dilation, which does not blur the image to fill the small gaps. Moreover, using the second view (Fig. 7.19) helps to reduce depth discontinuities caused by occlusion. Fig. 7.20 shows the rendered image using the reference image, which is rotated 16° about the camera optical axis.

Once the training images are acquired, we apply a key point detector on each image to extract and describe regions suitable for robust detection. In order to detect the satellite pose at a range of scales, scale-invariant features are learned during the training. Therefore, we employ the SIFT feature detector and descriptor [7] and represent each training image in a database of features. The SIFT descriptor is invariant to scale,

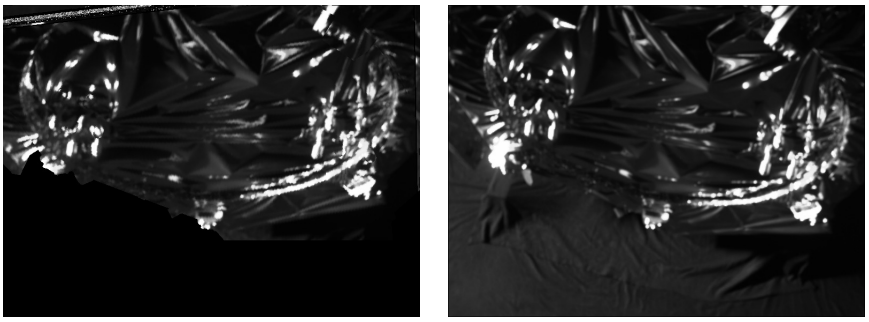


FIGURE 7.20: Image-based rendering for missing views during the training. The rendered image (left) is used for unreachable view points with the robot base.

rotation and to some extent to illumination. The scale-invariance of the SIFT is very important to handle the change of scale during matching with the query image, which may be far or close to the object. Each training image is represented by a text file containing the feature descriptor for each point and the location, orientation and scale of feature points. The vocabulary tree is built with feature descriptors, while scale and orientation can be in general used to check the geometric consistency.

Training in a Vocabulary Tree

The feature descriptors of all the training images are hierarchically clustered with the K-means clustering and built into the vocabulary tree of branch B_f and depth L_f according to [185]. At the root level each bag-of-feature is initially quantized to B_f branches. Each branch B_f is then sub-quantized sequentially and repeated until arriving the maximum depth (level). Each node within the tree is linked to its associated features as well as the images. Notice that the K-means clustering is continuously applied at each level of the tree to split the parent features and distribute them to B_f children nodes. We employ the square root kernel distance metric (Bhattacharyya distance) to measure the similarity between the feature descriptors as rootSIFT in [187]. The branch factor and the depth of the tree influence the retrieval result, however increasing these values does not necessarily improve the retrieval results.

Weighting of the nodes of the tree are applied to determine how each feature descriptor (word) from the quantized descriptors (code book) votes for each view. We employ a hierarchical weighting of the node, which depends on the number of features assigned to the node and its depth L_f. Therefore, the weight for each node and view relationship is

the normalized product of the feature histogram and level of the node. This weighting enables a depth-bounded search and is useful to reduce retrieval time. Related to the hierarchical weighting is a flat weighting, where only the leaf nodes are weighted with normalized histogram of features.

Retrieval and Scoring

Each feature of the query image searches for the nearest node on each level, applying the same distance metric used for the training. Thus, for each feature of the query image a path of selected nodes traverses the whole tree from the root node to the leaf node. The weights of the selected nodes and training image provide a similarity measure, which is used for retrieval of the best views for the query image.

The best match is retrieved based on the accumulated weights of each training image. We create a histogram with the list of weights of the related nodes to those selected by the bag of features from the training images. Finally images with the highest accumulated weight in the histogram are returned as the best match. In our experiment, we take the histogram of weights as a confidence measure in retrieving the corresponding image.

7.3.2 Feature Matching and and Pose Estimation

We extract and describe features of a query image with the Harris detector and the SIFT descriptor respectively. For each feature descriptor of the query image, we search the corresponding feature descriptor in the tree. After a successful retrieval of the corresponding image, we match features (Fig. 7.21). Notice that Harris feature detector is used for the online phase of the image matching and pose estimation due to its higher localization accuracy and robustness [188] as well as speed. Therefore, we employ Harris detector for both retrieved and query image while SIFT is used for the description of the region around the features. We remark that the SIFT features of the retrieved image, already stored and used for the training, could be used for online phase, however SIFT detector is less accurate and slower than Harris in our images. Speed up can be achieved by detecting new features at the expense of memory which requires to store Harris features of the training images.

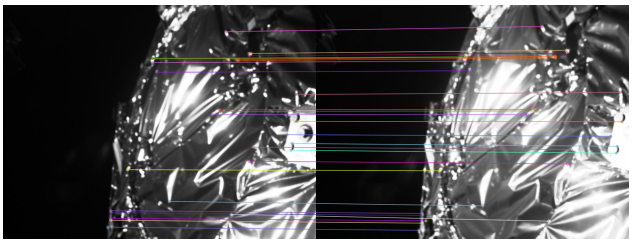


FIGURE 7.21: The feature matching for pose estimation. The features of the query image (right) are matched with the retrieved image (left) in the database.

Once we have the Harris features and SIFT descriptors of both query and retrieved image, we apply descriptor matching using a fast KD tree. Moreover, we compute the 3D points corresponding to the feature points of the training images, by back-projection using the depth map computed with the known pose and the CAD model. Then, given matched 2D features and the 3D points of the retrieved image, we can estimate the pose using the 2D-3D registration method (PnP problem). We adapt iterative efficient PnP [152] and RANSAC, because of the fact that after the matching not all the correspondences are correct. Therefore RANSAC is used to reject the false correspondences (outliers) while estimating the pose.

7.3.3 Evaluation

The appearance based pose detection method described in Section 7.3.1 is evaluated with real image sequences of gray image 780×582. We use feature descriptors, extracted from 2500 images and hierarchically quantized to build the vocabulary tree. The average processing time for image retrieval and pose estimation with Intel Xeon 2.8 GHz CPU is 0.8 second. The computation time depends on the number of keypoints and can be limited to improve the efficiency, but at the cost of reduced accuracy. The evaluation criteria is based on a ground truth obtained through the measurements of robot kinematics, a hand-eye and robot calibration. The experimental setup consists of a mock-up of the rear part of the TerraSAR-X satellite in full scale (Fig. 7.22), a camera mounted on a six DOF robot and the sun simulator. The TerraSAR-X satellite is placed in a fixed position and orientation, and the motion trajectories of the satellite are simulated by the six DOF robot motion. In this experiment, the sun simulator is placed in three main directions for the appearance training phase: $90°$, $30°$ and $-30°$ from the satellite mock-up. For the testing, the sun direction is rotated approximately $15°$ from the corresponding sun directions of the training phase, i.e. $75°$, $45°$ and $-15°$

FIGURE 7.22: Mock-up of the TerraSAR-X satellite (rear part with the launcher interface brackets) under normal illumination (color image) and space lighting (gray image).

respectively. A calibrated camera of resolution 780 × 582 pixel and a lens with focal length 6 mm was used to capture images. This choice of camera system ensures the required field of view within the close range of 1.5 m.

During training the vocabulary tree, a branch factor (width of the tree) of 6 and a level of 10 was selected. The goal of the experiment is to evaluate how accurately the pose of the satellite can be estimated. Here we present three image sequences that consider three scenarios; firstly we evaluate the pose estimation at two distances (1.3 m and 0.7 m) (Fig. 7.23 and 7.24), but with the same sun direction. This performance evaluation enables to determine how the algorithm behaves with different scale, in addition we study the effect of the sun direction at angle of 45°. The ground truth rotation and translation errors in most frames are 1° (Fig. 7.23) and 25 mm (Fig. 7.24) respectively, therefore the estimated pose can be obviously used to initialize a model-based tracking. The larger errors (spikes) in the plot are the frames which fail to be detected accurately. The pose estimation in this case results in a large error because of the inaccurately retrieved image which corresponds to a query image and insufficient valid feature correspondences. The detection failure is reported when the retrieval score is low (<45) or the number of correctly matched features is below the minimum (<4).

On the other hand, when the sun direction is about 75°, which is near to the normal (90°) to the line of sight of the camera, the distribution of the pose detection error (Fig. 7.25) is similar to Fig. 7.23 and 7.24 in spite of change of illumination. Similarly, the pose errors with the sun direction −15° (Fig. 7.26) indicate a comparable error distribution to that of the sun direction 75°. This is because of the fact that the sun direction for the appearance learning is close to that of the testing (differences of 15°), consequently the influence of illumination change is insignificant. However, it is necessary to train

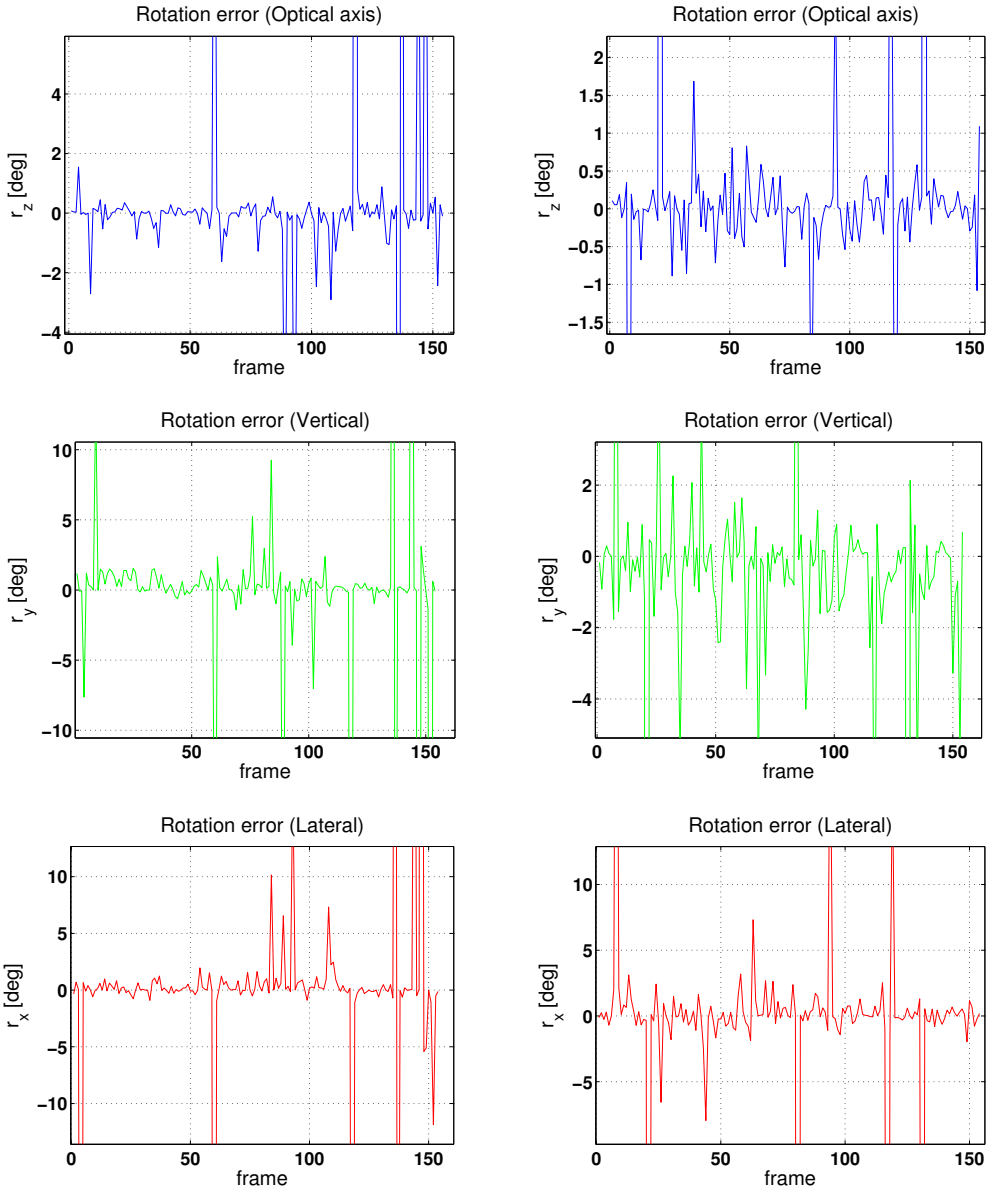


FIGURE 7.23: The rotation error in pose detection. The ground truth rotation error as evaluated 1.3 m (left column) and 0.7 m (right column).

the illumination at a sufficient interval in order to reduce the effect of the change of illumination on the pose detection. For example, in our experiment the pose detection fails completely when the test sequence was at the direction of 60°, i.e. the nearest training direction to the test sequence was 90° and 30°.

In general, sampling the sun direction in 3D space at an interval of 15° results in the desired accuracy in pose detection. However, such fine sampling creates hundreds of training sun directions, which may be unacceptable in terms of memory requirement to store respective features. Moreover, it increases the query time during retrieval. In

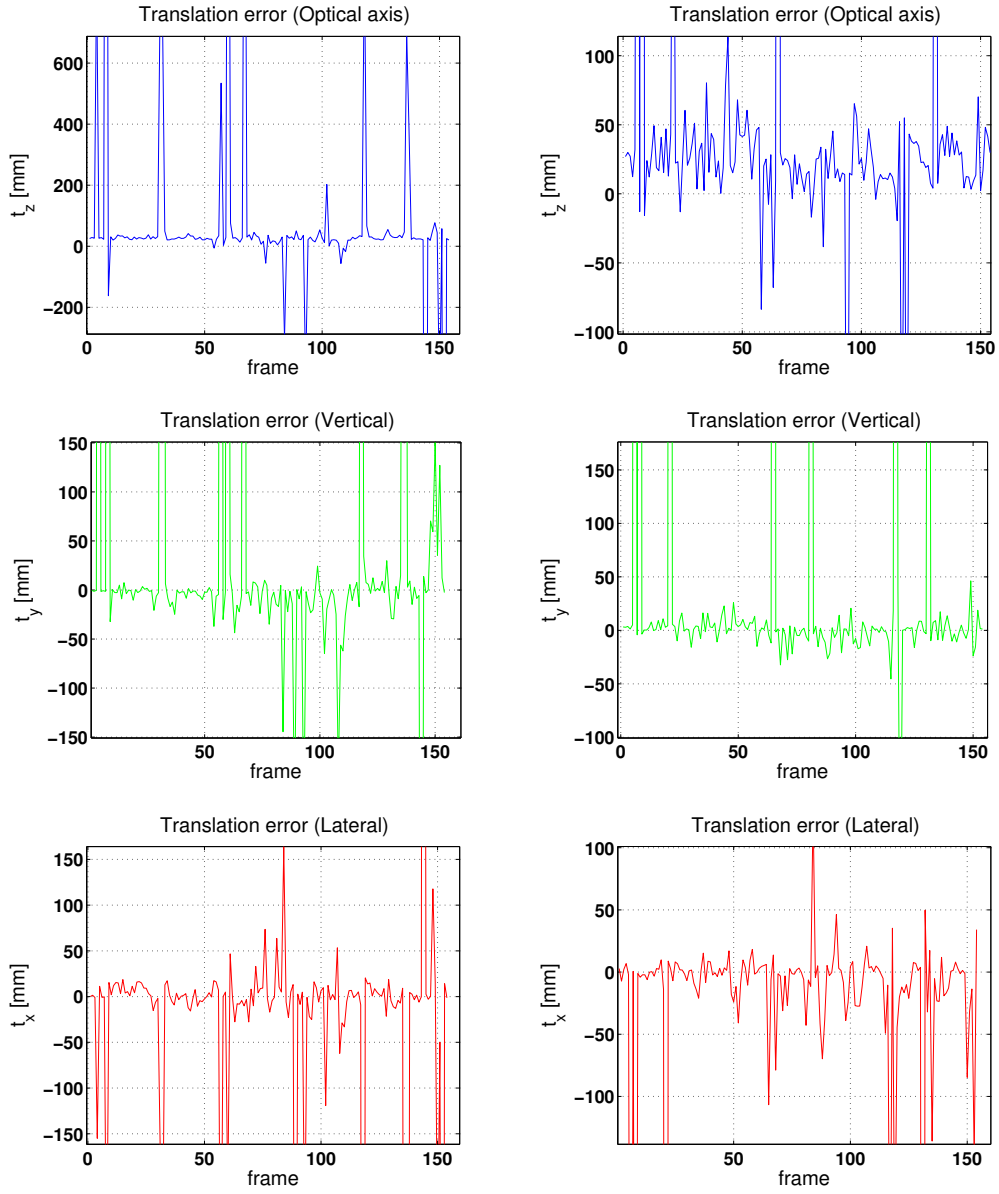


FIGURE 7.24: The translation error in pose detection. The ground truth translation error as evaluated at a distance of 1.3 m (left column) and 0.7 m (right column).

practice however, the sampling of the full sun angle is not necessary since the range of the sun direction will be known during a particular on-orbit servicing mission. Therefore, we restrict the possible directions of sun to a certain range provided by the on-orbit servicing mission. This in turn reduces the memory requirement for storing database features and query time.

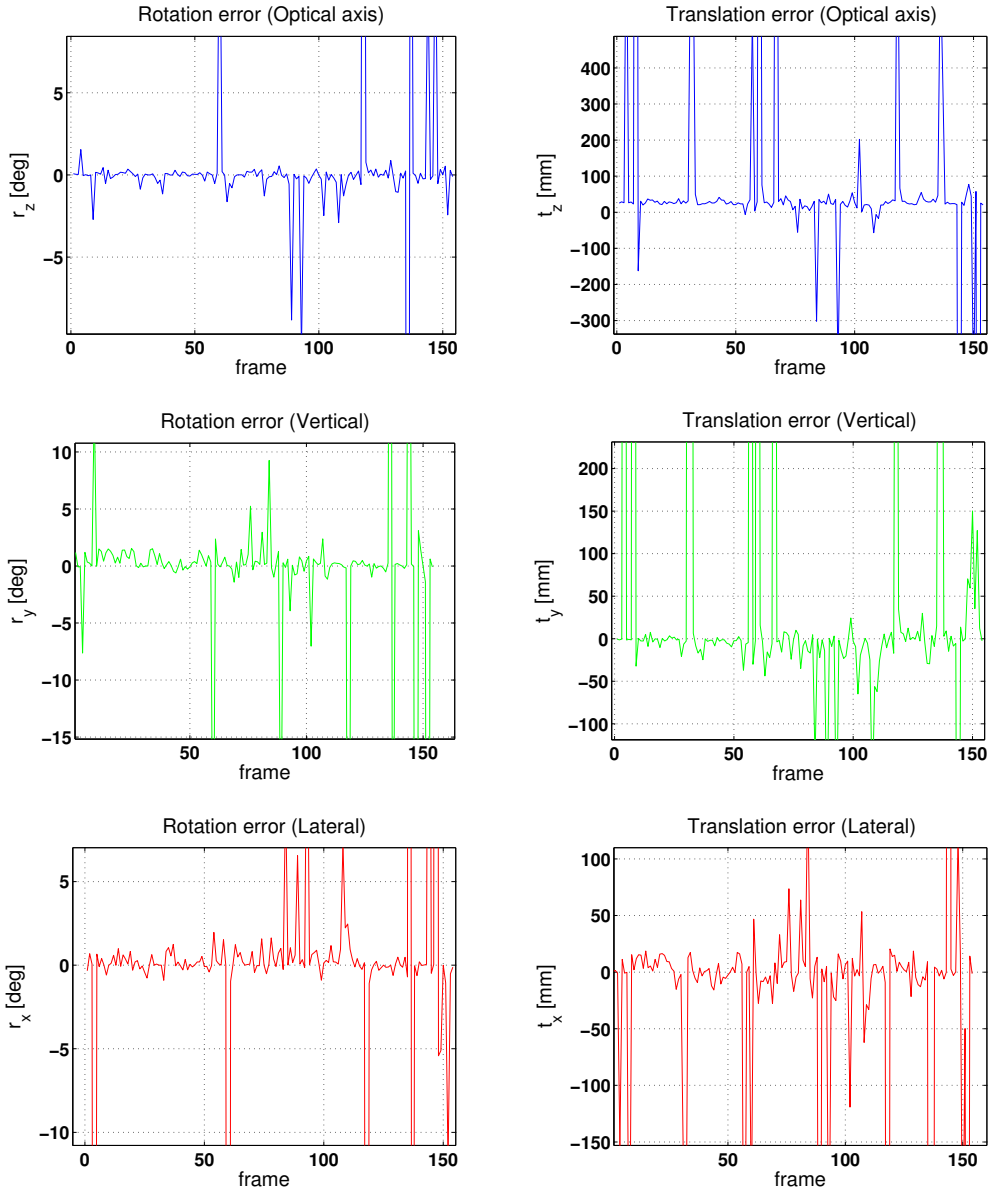


FIGURE 7.25: The accuracy of the pose detection compared to ground truth with the sun direction of 75°. The ground truth rotation (left column) and translation (right column) error at a distance of 1.3 m.

7.3.4 Summary

The model-based tracking relies on the last pose to estimate the current pose using local optimization. Therefore, it is essential to detect the satellite in an image and estimate its pose. The initial pose estimation, i.e. pose detection is generally a challenging problem in computer vision and image processing. The problem of pose detection is more difficult under space lighting. The geometry of the object also determines the performance of the detection. On one hand, the pose of an object with several distinctive edges is

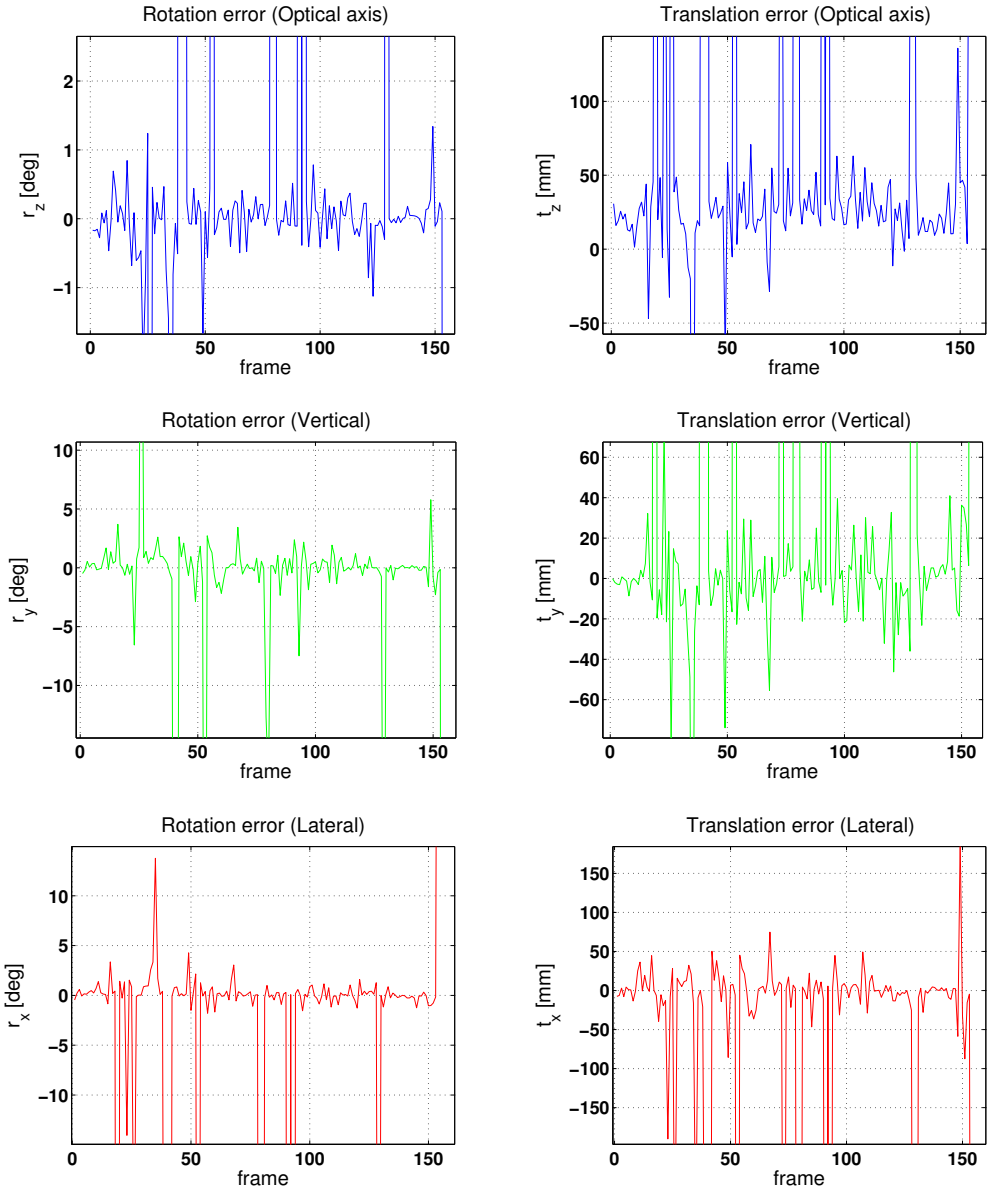


FIGURE 7.26: The accuracy of the pose detection compared to ground truth with the sun direction of $-15°$. The ground truth rotation (left column) and translation (right column) error at a distance of 1.3 m.

estimated with edge-based methods. On the other hand, feature-based methods are used to initialize textured objects. However, in certain cases the region of interest of the satellite may consist of highly reflective multilayer insulation (MLI) and dominates useful features such as edges. The MLI poses difficulty in edge-based detection because of reflection and virtual unmodeled edges. In this thesis, we address the pose detection of the TerraSAR-X satellite fully covered with an MLI in close-range. The rear part of the TerraSAR-X satellite consists of launcher interface brackets, which are suitable for grasping during robotic manipulation.

The pose detection is based on the appearance learning of the full scale mock-up of the TerraSAR-X satellite. We use several images under various view points and sun direction for the training of a vocabulary tree. The vocabulary tree is a tree structure used to represent the training images by quantizing respective feature descriptors with the hierarchical K-means clustering. We employ a 3D image warping to synthesize missing images because of the limited workspace of the robot, which simulates the motion trajectory of the satellite. After successful retrieval of the corresponding image to the query image, we match feature descriptors with the fast KD tree. The 3D points corresponding to the training features are computed using the depth map; the depth map is in turn obtained by rendering the 3D model with the Z-buffer. The pose is estimated from the correspondence based on iterative 3D-2D registration and RANSAC.

We validate the proposed pose detection method with a challenging space lighting condition. The evaluation criteria is based on the accuracy of the estimated poses with respect to the ground truth. The pose detection is evaluated at three sun directions, which deviate 15° from the sun direction of the training, and indicates 90% detection rate. Moreover, we observe that as the disparity between training and testing sun directions becomes greater than 15°, the detection rate decreases significantly.

Chapter 8

Conclusion

The increasing number of space debris as well as malfunctioning satellites in orbit has recently attracted attention of major space agencies to mitigate the congestion. An on-orbit servicing is a mitigation technique of orbital debris which has congested the precious orbits and has endangered satellites. The on-orbit servicing task includes repairing, maintenance, refueling and de-orbiting of the malfunctioning client satellite. Motion estimation and tracking of the malfunctioning satellite is required for the approach and capture phase of the on-orbit servicing. This thesis addresses the problem of vision-based motion estimation and tracking of a client satellite for the on-orbit servicing. We rely on camera images to estimate the position as well as orientation of the client satellite pursuing novel approaches to deal with various image properties at various ranges under harsh lighting conditions and highly reflective multilayer insulation (MLI).

LIDAR based measurements are relatively robust to space lighting and have been often used for formation flying and rendezvous to ISS. However, the LIDAR has high mass and volume, and consumes more power; this is against the stringent requirements of a satellite design and increases the costs significantly. In this regard, a camera based localization is an inexpensive solution for an on-orbit servicing mission, although the camera sensor is highly sensitive to space lighting and the reflective surface of the multilayer insulation. The vision-based localization of a highly reflective satellite is still a difficult problem in spite of a number of efforts in current literature. The camera based localization of a satellite with a predominantly MLI surface is barely addressed for the on-orbit servicing

from a practical point of view and this thesis contributes in this context in three main areas.

Among our main contributions [162, 163, 166, 167, 173] is the point-wise predictive model used for tracking and prediction of 3D points associated to each visual feature in the absence of a CAD model of the satellite, discussed in Section 8.1. The second contribution of the thesis is a novel hierarchical position localization of nearly malfunctioning satellites relying on a model-based foreground and background segmentation, and edge tracking (Section 8.2.1). Related to the model-based contribution is a real time pose tracking in six degrees of freedom by representing the CAD model into parts, and adaptive search distance along the edge normals to disambiguate a set of edges (Section 8.2.2). The third contribution is a model-based appearance learning for pose detection of a satellite fully covered with MLI using a vocabulary tree (Section 8.2.3). Our contributions are effectively demonstrated experimentally with realistic images and full scale satellite mock-ups using the most realistic European rendezvous and proximity operations simulator at DLR. The experimental findings are chapter specific and are summarized according to the sequence of the Chapters presented in the thesis. Finally we present future perspectives.

8.1 Model-free Motion Estimation

The geometric model of a malfunctioning client satellite may not exist or is not any more usable. In this case, we can estimate only relative frame-to-frame motion of the satellite in a camera coordinate frame. The motion estimation and tracking in the absence of the CAD model of the satellite poses challenges because of the reflection of the MLI and lack of geometric a priori information. Stereo based methods provide accurate solutions to this problem if sufficient stable features are available. The MLI of the satellite surface wrinkles provide sparse features for tracking. However, these features may disappear shortly and new features appear because of reflection and under- or over-illumination.

This thesis presents a novel motion prediction method for 3D point features, which is used in case of temporary missing feature measurements (Chapter 5). The algorithm builds up on sparse stereo matching, visual feature tracking and motion prediction. The motion of each 3D point associated with a visual feature is predicted based on a

point-wise kinematic model. The point-wise kinematic motion model is derived from the rigid-body velocity. The velocity is in turn estimated from an image motion and depth. We employ the iterated Kalman filter to estimate the motion of each 3D point, predict according to this motion model and update with the tracked low-level features.

The proposed method is demonstrated on real images of a client satellite in the presence of at least minimum features for pose estimation. In the absence of minimum features, this thesis shows that the motion of the client satellite can be estimated using a novel motion prediction method based on the motion model derived from the rigid-body velocity. The motion prediction relies on the last estimated velocity, which is used for the next few frames. However, the prediction capability of the motion model is limited and cannot handle long-term motion prediction, which is beyond the scope of this thesis.

8.2 Model-based Pose Estimation

In many cases, the CAD model of client satellites can be obtained from the satellite designers and manufacturers. The geometric model of the satellite provides a priori information about the shape of the satellite. Incorporating the shape model into the pose estimation greatly helps to avoid drift and provide poses in absolute Cartesian frame. In this thesis, we address two practical problems related to model-based tracking under harsh space lighting condition. Firstly, we present a novel position detection and tracking method of a partially cooperative satellite in the presence of a simple model (Section 8.2.1). In this case, the client satellite has functional attitude sensors (partially cooperative). Secondly, we provided position and orientation (pose) tracking of a tumbling satellite in the presence of the CAD model (Section 8.2.2).

8.2.1 Position Localization of a Partially Cooperative Satellite

Localization of a partially cooperative satellite is of great benefit for on-orbit servicing because of the on-board attitude measurement sensors. The attitude sensors of the client can be used to align with attitude of the servicer so that only position estimation is required during the approach phase. However, the localization of the client satellite with MLI is yet challenging because of the specular reflection and direct sun lighting. In particular the problem of position localization in various ranges poses a number of

difficulties. For example, the image properties at mid- and close-range varies significantly because of reflective MLI, under- and over-illumination. To address this problem, the thesis introduces a hierarchical localization method that takes account image properties at various ranges (Chapter 6). The hierarchical tracking is based on a model-based foreground and background segmentation in the mid-range and edge-based tracking at close-range. The main contribution lies in the mid-range that uses the model of the silhouette contour of the satellite and image histograms to simultaneously estimate the position, and segment foreground from background. At close-range, we use rather universal primitive features of the satellite such as the engine nozzle for edge-based localization. The close-range approach is a critical phase and requires higher accuracy and robustness in tracking. Therefore, the performance evaluation takes into account the effect of image compression, sunlight direction and level of target illumination system on the tracking. In case of tracking failure due to unpredicted motion and over-saturation of the image at close-range, the thesis also provides a novel nozzle detection method that integrates RANSAC and absolute geometric error fitting into the model based stereo matching (Chapter 6, Section 6.3.1).

We have performed an extensive evaluation of the position localization system using a realistic proximity operations simulator that reproduces the satellite surface properties and space lighting. The position tracking based on a foreground and background segmentation at mid-range is shown to be accurate and robust even with compressed images and poor lighting. The thesis pays particular attention to the evaluation of the critical close-range approach. The tracking performance depends on the direction of the sun and the satellite with respect to the camera axis. The empirical results indicate that when the sun angle approaches normal to the client satellite and the line of sight of the camera, an additional source of light should be mounted on the servicer satellite for reliable tracking of the nozzle because of the poor illumination and shadowing. In contrast, the level of the target illumination is also critical and should be reduced or removed as the sun angle drifts significantly from the normal. These findings are strictly concerned with attitude controlled client satellites in frontal position, and does not generalize to tumbling satellites, because the change of view during tumbling changes illumination conditions (discussed in Section 8.2.2).

8.2.2 Pose Tracking of a Tumbling Satellite

A client satellite may undergo spinning and tumbling motion about any of its axes freely because of the failure of the control system. On-orbit servicing of a tumbling satellite is particularly challenging because of the arbitrary motion and reflection of the MLI. Mostly we have the knowledge of the mesh model of the satellite. The CAD model of the satellite is of great benefit for an accurate and robust tracking. In the presence of the full CAD model, the pose estimation is achieved by aligning the visible model and image edges along the edge normal at a given pose hypothesis, provided by the last pose. However, matching edges by 1D search along the edge normal from far distance creates ambiguities for closely located image edges. This thesis provides a novel method to solve edge ambiguities by introducing an adaptive search length along the edge normal which depends on the object distance from the camera, object size and properly selected focal length of the lens (Chapter 7).

We have evaluated the pose tracking in six DOF based on the realistic proximity operations hardware facility as well as photo-realistic images obtained from rendering. The adaptive search length is demonstrated to be more effective to ambiguous edges. Similar to the partially cooperative satellite, the sun direction significantly influences the tracking performance. The experimental results also indicate that the performance of the tracking depends on the type of motion (rotation) of the client satellite. It implies that the performance of the tracking is not linear in terms of the sun angle, and rather depends on the type of motion (in-plane, out-of-plane, nutation angles).

8.2.3 Pose Detection of a Satellite with a Dominant MLI Surface

The tracking based on a local optimization requires an initial pose estimation. The problem of pose detection is generally more difficult because of the large search space and several local optima. In particular, the MLI poses difficulties with edge-based detection because of reflection and virtual edges associated with specular reflection. In this thesis, we present a method for the pose detection of a satellite fully covered with an MLI at close range. The pose detection is based on the appearance learning of a full scale mock-up of the TerraSAR-X satellite. Real representative images under various view points and sun direction are used to train a vocabulary tree. The 3D image warping is

employed to obtain some of the missing views due to limited workspace of the six degrees of freedom robot that carries camera to sample view points. The vocabulary tree is used to represent the training images by quantizing respective feature descriptors with the hierarchical K-means clustering. After successful retrieval of the corresponding image to the query image, the 3D points corresponding to the training features are computed by rendering the 3D model. The 3D pose is then estimated from the 3D-2D correspondence.

The pose detection method has been evaluated with a challenging space lighting condition, based on the ground truth obtained from the calibrated robot and camera system. The experimental results indicate that the pose detection rate increases significantly as the sun direction of test image is closer to that of the training images. This implies that at the expense of a memory for storage of several image features under more illumination direction, the accuracy and the robustness of pose detection can be improved. The main assumption in this approach is the fact that the training with photo-realistic images can be obtained by re-building a mock-up of the client satellite. Notice that only the outer structure of the satellite would be used as a mock-up, hence it should not be too expensive. In case of large satellites, beyond the robot carrying capacity, a scaled mock-up can be still used for the training of the pose detection algorithm.

The thesis has demonstrated that inexpensive camera based localization methods can be effectively employed for the approach and capture phase of on-orbit servicing mission. The thesis attempts to focus on real typical problems that occur in on-orbit servicing missions, related to reflective multilayer insulation (MLI), sun lighting and satellite motion. A target illumination system is required under very poor illumination condition, increasing overall mass and power consumption of the camera based localization. Nevertheless, the mass, volume and power consumption of the whole camera system can be designed lower than that of a typical LIDAR based system. The new developments in both camera and LIDAR technologies in terms of mass, volume, power and sensitivity to space lighting may continue to compete for space applications. In this regard, plenoptic cameras provide better depth of field independent of aperture than that of conventional cameras, and could be promising for future developments.

8.3 Prospective Questions

The thesis has addressed practical problems of visual tracking for the approach and manipulation of a satellite in the context of an on-orbit servicing. It focused on the specific application scenarios that encounter in real on-orbit servicing missions. In particular, it addressed the challenging space lighting and satellites covered fully with a multilayer insulation under certain assumptions. The main assumptions include low tumbling rate of the non-cooperative client satellite, smooth maneuver and a calibrated camera system. In spite of these achievements within the mentioned assumptions, potential improvements can be performed.

In the absence of a CAD model, the 3D point tracking relies on low-level visual features which are tracked based on optical flow. While the optical flow provides accurate localization of features, it relies on the classical constant brightness assumption. This assumption holds for slowly tumbling satellite, where the frame-to-frame motion is small and a sufficient number of features satisfy the constraint. However, for satellite tumbling at a higher rate the change of illumination is significant and the assumption does not hold any more. The method could be improved by more illumination invariant feature descriptors such as SIFT and SURF and metric-based matching, but at the expense of computation time.

The localization of a partially cooperative satellite is one of the most reliable approaches for rendezvous and on-orbit servicing. The thesis addressed the localization under difficult space lighting conditions and highly reflective MLI by carefully selecting image features at various ranges and hierarchically integrating robust tracking and detection methods. The satellite dynamic parameters such as inertia matrix of such a client satellite may be known, and therefore could be used for accurate prediction of the motion. These would help the localization system in case of sudden unexpected motion, and even under bad illumination conditions. Thus, the future research could be directed into the exploitation of the known dynamic parameters in the localization system which might be profitable for vision-based rendezvous of a partially cooperative satellite.

The thesis has demonstrated the pose detection based on appearance learning assuming the mock-up of the client satellite can be precisely reproduced. The real camera images taken from the mock-up were used for the training view points at various sun direction.

Instead of using the mock-up, photo-realistic images should be obtained by rendering the CAD model of the satellite. However, the current technology does not allow to accurately model MLI. Therefore, a possible area in future investigations includes modeling of the reflective property of the MLI for more accurate photo-realistic rendering. The accurate modeling of MLI in computer graphics opens vast opportunities in research for better tracking and detection algorithms.

Appendix A

Space Rendezvous and Orbital Debris

A.1 Satellite Rendezvous: Brief History

A space rendezvous is an orbital maneuver in which two spacerafts approach to a very close distance in the same orbit. The rendezvous may be followed by docking or berthing, which brings the two spacecrafts to physical contact [2]. Table A.1 lists some of the rendezvous and docking missions.

TABLE A.1: Brief history of satellite rendezvous missions.

Year	Rendezvous mission
1965	Gemini VII, NASA's manned space mission to effect rendezvous and docking and to demonstrate extravehicular activity and other technologies.
1967	The first autonomous rendezvous, by Soviet vehicles Cosmos 186 and 187
1969	The rendezvous and docking with an exchange of crew by Soyuz 4 and 5
1971-1999	Soyuz manned and progress unmanned spacecrafts of Russia performed autonomous docking to the first space station Salyut and Mir
1975	International docking mission (Russia,USA) between the Apollo and the Soyuz spacecrafts
1984	The first US space shuttle, designed for space transportation system (STS), repaired solarMax Satellite and provide crew exchange missions with Mir space station in 1990s until 1999

Year	Rendezvous mission
1993-2009	Manned satellite servicing of Hubble space telescope involving space shuttle to repair or replace optical instruments.
1997	NASDA robotic mission aimed to carry out several experiments related to rendezvous, docking and space robotics
1998	The space shuttle mission to the international space station
2007	DARPA mission to demonstrate several satellite servicing operations and technologies including rendezvous, proximity operations, station keeping, capture and docking using ASTRO and NEXTsat satellites
2008	The Automated Vehicle (ATV) of ESA has successfully delivered supplies to ISS with fully autonomous docking capabilities
2009	Japanese H-II Transfer vehicle (HTV) delivered supplies to ISS with partially autonomous docking
2010	PRISMA mission by Swedish space corporation for autonomous rendezvous and proximity operation between Mango and Tango satellites
2012	The first privately owned commercial craft by SpaceX supplied to the ISS fully autonomously with its Dragon spacecraft

A.2 Orbital Debris

Orbital debris is man-made object in orbit around the Earth which is no longer used. Such debris includes malfunctioning spacecraft, mission related debris, fragmentation debris and abandoned launch vehicle stages such as rocket bodies. The debris is increasing, occupying useful orbits and endangering satellites. The quantitative data presented by NASA, shown in Fig. A.1 indicates the debris will drastically increase and congest the Low Earth and Geostationary Orbits in future. In particular, most orbital debris resides within a few thousand km from the Earth's surface. The amount varies significantly with altitude. The highest concentrations of orbital debris are found around 800 km. In Low Earth Orbit, the debris orbits at speed of up to 8 km/s, which involves significant energy during collision [1].

[1] http://orbitaldebris.jsc.nasa.gov/

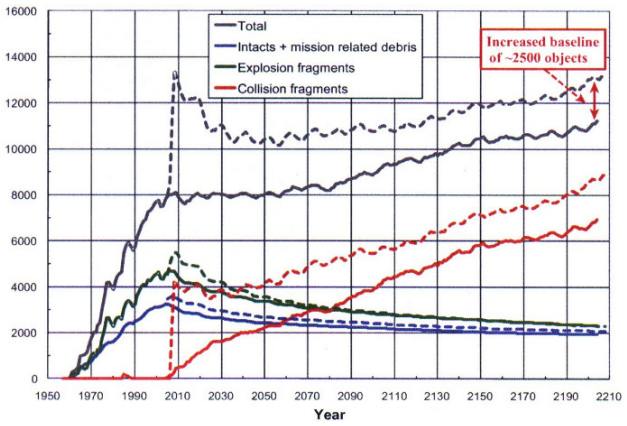


FIGURE A.1: The trend of space debris, current and the predicted for the next few years. Image courtesy of NASA.

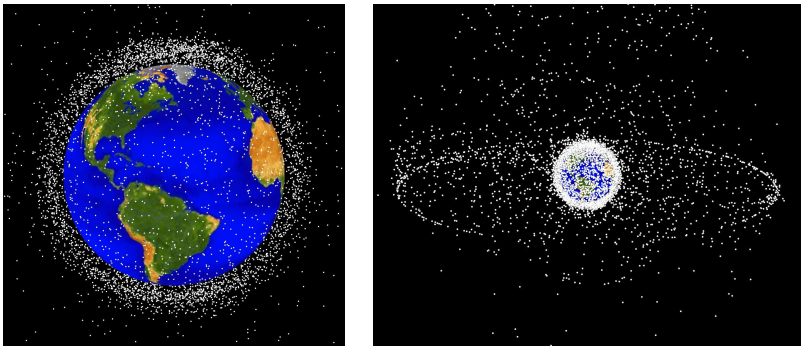


FIGURE A.2: Orbital debris in LEO and GEO. The most concentration of debris in LEO (left) and debris in geosynchronous region near GEO (Right). Image courtesy of NASA.

Appendix B

Rotation Parametrization

There exist well known parametrizations for the special orthogonal group $SO(3)$ of rotations in $\mathbb{R}^3$. The SO(3) group consists of orthogonal matrices R_3, with $R_3^T R_3 = I_3$ and $\det(R_3) = 1$. The orthogonality constraint helps reduce the degrees of freedom of a rotation matrix with respect to the total number of parameters, i.e. the number of parameters in rotation matrix is 9 (Equation (B.1)), while in SO(3) we have only 3 free parameters. However, parametrization in SO(3) is complex because of several possible solutions with different properties. There exist three common representation of a rotation matrix: Euler angles, axis-angle and quaternion.

$$R_3 = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{B.1}$$

B.1 Euler angle

Euler angle is the most intuitive representation of a rotation matrix. However, the representation has singular configurations called gimbal lock. The rotation matrix is expressed as a cascade of three elementary rotations around each axis of Cartesian frame, for example rotation $R(\phi, \theta, \alpha)$ in the order x, y, and z is represented by

$$R(\phi, \theta, \alpha) = R_x(\phi) R_y(\theta) R_z(\alpha) \tag{B.2}$$

where rotation about each axis is

$$R_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\phi & -sin\phi \\ 0 & sin\phi & cos\phi \end{bmatrix}, R_y(\theta) = \begin{bmatrix} cos\theta & 0 & -sin\theta \\ 0 & 1 & 0 \\ sin\theta & 0 & cos\theta \end{bmatrix} and \tag{B.3}$$

$$R_z(\alpha) = \begin{bmatrix} cos\alpha & -sin\alpha & 0 \\ sin\alpha & cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{B.4}$$

The singular configuration arises while computing the inverse problem, i.e. the pose parameters from the rotation matrix R. For example, for $\theta = \frac{\pi}{2}$ any pair $\phi = \alpha$ provides the same rotation matrix, hence there exist infinite solutions.

B.2 Axis-angle

Rotation is parametrized by the pair $(\mathbf{u}, \theta)$, where θ is a rotation angle and $\mathbf{u} = [u_1 \quad u_2 \quad u_3]^T$ rotation axis. The three-parameter representation is given by the rotation vector $\rho = \mathbf{u}\theta$. Any rotation matrix R has a unit eigenvalue corresponding to a given eigenvector $\mathbf{u}$ such that $R\mathbf{u} = \mathbf{u}$. R is given by the Rodrigues rotation formula:

$$R(\rho) = I + Usin\theta + U^2(1 - cos\theta) \tag{B.5}$$

where U=$[\mathbf{u}]_\times$ is a skew-symmetric matrix given by

$$U = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & o & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}. \tag{B.6}$$

For small angle $\theta \to 0$, the rotation is approximated by

$$R(\rho) = I + [\rho]_\times \tag{B.7}$$

and corresponds to the Lie algebra representation of the tangent space SO(3). The advantage of the angle-axis representation is that the inverse problem $R \to \rho$ always

provides a single solution, given by

$$\theta = cos^{-1}\left(\frac{trace(R)-1}{2}\right), \quad \mathbf{u} = \frac{1}{2sin\theta}[r_{32}-r_{23} \quad r_{13}-r_{31} \quad r_{21}-r_{12}]^T \tag{B.8}$$

except for $\theta = \pi$ which has two solutions, $\pm\rho$. Obviously, for θ=0, ρ=0.

B.3 Quaternion

Quaternions are hypercomplex numbers, given by $(q_s + q_x\mathbf{i} + q_y\mathbf{j} + q_z\mathbf{k})$, where $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk}$ =-1. The 4-tuple quaternion has scalar q_s and vector $\mathbf{q} = (q_x, q_y, q_z)$ components. The unit quaternion is defined by introducing the Euclidean norm $\|q\|^2$=$q_s^2 + q_x^2 + q_y^2 + q_z^2$ and $\|q\|$=1. The unit quaternions are commonly used to parametrize 3D rotations, and are related to rotations in SO(3).

Let $\mathbf{r}$ be a vector, and $(0, \mathbf{r})$ be the imaginary quaternion associated with $\mathbf{r}$. The unit quaternion q is given by $q = (cos(\frac{\theta}{2}), \mathbf{q}sin(\frac{\theta}{2}))$, with $\|\mathbf{q}\| = 1$ for some $(\theta, \mathbf{q})$,

and we compute the product

$$\mathbf{r}' = q \bigotimes \mathbf{r} \bigotimes \bar{q} \tag{B.9}$$

where $\bigotimes$ represents internal product (Hamilton product) and $\bar{q}$ is the conjugate of quaternion q. This operation is equivalent to rotating $\mathbf{r}$ into $\mathbf{r}'$ around the angle-axis $(\theta, \mathbf{q})$. The unit quaternion is equivalent to a rigid rotation matrix,

$$R(q) = \begin{bmatrix} q_s^2 + q_x^2 - q_y^2 - q_z^2 & 2q_xq_y - 2q_sq_z & 2q_sq_y + 2q_xq_z \\ 2q_sq_z + 2q_xq_y & q_s^2 - q_x^2 + q_y^2 - q_z^2 & 2q_yq_z - 2q_sq_x \\ 2q_xq_z - 2q_sq_y & 2q_sq_x + 2q_yq_z & q_s^2 - q_x^2 - q_y^2 + q_z^2 \end{bmatrix} \tag{B.10}$$

and numerically stable for pose estimation. It is a non-singular representation of 3D rotation, with a one-to-two correspondence between rotation and a unit quaternion $(R \leftrightarrow \pm q)$. However, it introduces one more parameter, increasing dimension and the constraint $\|q\|$=1 must be enforced to prevent infinite solutions.

Bibliography

[1] Toralf Boge and Ou Ma. Using advanced industrial robotics for spacecraft rendezvous and docking simulation. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 1–4, May 2011.

[2] Wigbert Fehse. Automated rendezvous and docking of spacecraft. *Cambridge University Press*, 2003.

[3] F. Sellmaier, T. Boge, J. Spurmann, S. Gully, T. Rupp, and F. Hubber. On-orbit servicing missions: Challenges and solutions for spacecraft operations. In *Proceedings of the AIAA SpaceOps conference*, Huntsville, Alabama, April 2010.

[4] Tom Drummond and Roberto Cipolla. Real-time visual tracking of complex structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7): 932–946, 2002.

[5] Andrew I. Comport, Danica Kragic, Éric Marchand, and François Chaumette. Robust real-time visual tracking: Comparison, theoretical analysis and performance evaluation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 2841–2846, April 2005.

[6] Jianbo Shi and Carlo Tomasi. Good features to track. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994.

[7] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2(60):91–110, 2004.

[8] Hernán Badino. A robust approach for ego-motion estimation using a mobile stereo platform. In *Proceedings of the 1st International Conference on Complex Motion*, pages 198–208, 2007.

[9] Michael W. Walker and Lejun Shao. Estimating 3-d location parameters using dual number quaternions. *CVGIP : Image Understanding*, 54(3):358–367, 1991.

[10] Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transaction Pattern Analysis and Machine Intelligence*, 13(4):376–380, 1991.

[11] Bogdan Matei and Peter Meer. Optimal rigid motion estimation and performance evaluation with bootstrap. In *Proceedings of IEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 339–345, 1999.

[12] Shai Segal, Avishy Carmi, and Pini Gurfil. Vision-based relative state estimation of non-cooperative spacecraft under modeling uncertainity. In *Proceedings of IEEE Aerospace Conference*, volume 1, pages 1–8, 2011.

[13] Jed M. Kelsey, Jeffrey Byrne, Martin Cosgrove, Sanjeev Seereeram, and Raman K. Mehra. Vision-based relative pose estimation for autonomous rendezvous and docking. In *Proceedings of IEEE Aerospace Conference*, Big Sky, MT, 2006.

[14] Antoine Petit, Eric Marchand, and Keyvan Kanani. A robust model-based tracker combining geometrical and color edge information. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3719–3724, 2013.

[15] Andrew I. Comport, Eric Marchand, and François Chaumette. Robust model-based tracking for robot vision. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and System*, volume 1, pages 692–697, 2004.

[16] Giorgio Panin, Erwin Roth, and Alois Knoll. Robust contour-based object tracking integrating color and edge likelihoods. In *Proceedings of the Vision, Modeling, and Visualization Conference*, pages 227–234, October 2008.

[17] P. Bodin, R. Noteborn, R. Larsson, S. D'Amico, J.S. Ardaens, M. Delpech, and J.C. Berges. The prisma formation flying demonstrator: Overview and conclusion from the nominal mission. In *Proceedings of the 35th annual AAS Guidance and Control Conference*, Colorado, USA, February 2012.

[18] Noriyasu Inaba and Mitsushige Oda. Autonomous satellite capture by a space robot: world first on-orbit experiment on a japanese robot satellite ETS-VII. In

Proceedings of IEEE International Conference on Robotics and Automation, volume 2, pages 1169–1174, 2000.

[19] James Shoemaker and Melissa Wright. Orbital express space operations architecture program. In *Proceedings of International Symposium on Aerospace and Defense Sensing Simulation and Controls (Aerosense)*, pages 1–9, 2003.

[20] Clemens Kaiser, Fredrik Sjöberg, Juan Manuel Delcura, and Baard Eilertsen. SMART-OLEV—An orbital life extension vehicle for servicing commercial spacecrafts in GEO. *Journal of Acta Astronautica*, 63(1–4):400–410, July 2008.

[21] Manny R. Leinz, Chih-Tsai Chen, Peter Scott, William Gaumer, Peter Sabasteanski, and Michael Beaven. Modeling, simulation, testing, and verification of the orbital express autonomous rendezvous and capture sensor system (ARCSS). In *Proceedings of SPIE on Sensors and Systems for Space Applications II*, volume 6958, pages 69580C–69580C–13, 2008.

[22] Andrew J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proceedings pf 9th IEEE International Conference on Computer Vision*, pages 1403–1410, 2003.

[23] Andrew J. Davison, Ian D. Reid, Nicholas D. Molton, and Olivier Stasse. MonoSLAM: real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067, June 2007.

[24] Javier Civera, Oscar G. Grasa, Andrew J. Davison, and J. M. M. Montiel. 1-point RANSAC for extended kalman filtering: Application to real-time structure from motion and visual odometry. *Journal of Field Robotics*, 27(5):609–631, 2010.

[25] Georg Klein and David Murray. Parallel tracking and mapping for small AR workspaces. In *Proceedings of 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, page 225–234, 2007.

[26] Richard A. Newcombe, Steven J. Lovegrove, and Andrew J. Davison. DTAM: dense tracking and mapping in real-time. In *Proceedings of IEEE International Conference on Computer Vision*, page 2320–2327, 2011.

[27] Jakob Enge, Jurgen Sturm, and Daniel Cremers. Semi-dense visual odometry for a monocular camera. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1449 –1456, 2013.

[28] Christian Forster, Matia Pizzoli, and David Scaramuzza. SVO: fast semi-direct monocular visual odometry. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 15–22, 2014.

[29] Clark F. Olson, Larry H. Matthies, Marcel Schoppers, and Mark W. Maimone. Rover navigation using stereo ego-motion. *Journal of Robotics and Autonomous Systems*, 43(4):215–229, June 2003.

[30] Annalisa Milella and Roland Siegwart. Stereo-based ego-motion estimation using pixel tracking and iterative closest point. In *Proceedings of IEEE International Conference on Computer Vision Systems*, page 21, 2006.

[31] F. Souvannavong, C. Lemaréchal, L. Rastel, M. Maurette, and France Magellium. Vision-based motion estimation for the ExoMars rover. In *Proceedings of International Symposium on Artificial Intelligence Robotics and Automation in Space (i-SAIRAS)*, Sapporo, Japan, 2010.

[32] P. Saeedi, D. Lawrence, and D. Lowe. 3d motion tracking of a mobile robot in a natural environment. In *Proceedings of IEEE International Conference on Robotics and Automation*, volume 2, pages 1682–1687, San Francisco, CA, 2000.

[33] Andrew I. Comport, E. Malis, and P. Rives. Real-time quadrifocal visual odometery. *International Journal of Robotics Research*, 29(2-3):245–266, 2010.

[34] Giorgio Panin and Nassir W. Oumer. Ego-motion estimation using rectified stereo and bilateral transfer function. In *Proceedings of Advances in Visual Computing*, pages 458–469. 2012.

[35] P.H.S Torr and A. Zisserman. Feature based methods for structure and motion estimation. In *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, pages 278–294, 2000.

[36] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Proceedings of Fourth Alvey Vision Conference*, pages 147–151, 1988.

[37] S.M. Smith and J.M. Brady. Susan- a new approach to a low level image processing. 23(1):1207–1212, 1997.

[38] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In *Proceedings of European Conference on Computer Vision*, pages 430–443. 2006.

[39] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (SURF). *Journal of Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[40] Motilal Agrawal, Kurt Konolige, and Morten Rufus Blas. Censure: Center surround extremas for realtime feature detection and matching. In *Proceedings of European Conference on Computer Vision*, pages 102–115. 2008.

[41] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. In *Proceedings of European Conference on Computer Vision*, pages 778–792. 2010.

[42] Stefan Leutenegger, Margarita Chli, and Roland Yves Siegwart. BRISK: binary robust invariant scalable keypoints. In *Proceedings of IEEE International Conference on Computer Vision*, pages 2548–2555, 2011.

[43] Alexandre Alahi, Raphael Ortiz, and Pierre Vandergheynst. Freak: Fast retina keypoint. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 510–517, 2012.

[44] K.S. Arun, T.S. Huang, and S.D. Blosten. Least-squares fitting of two 3d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–700, 1987.

[45] Berthold K.P. Horn. A closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America*, 4(4):629–642, 1987.

[46] D. Demirdjian and T. Darrell. Motion estimation from disparity images. In *Proceedings of IEEE International Conference on Computer Vision*, volume 1, pages 213 – 218, 2001.

[47] Louis-Philippe Morency and Trevor Darrell. Stereo tracking using icp and normal flow constraint. In *Proceedings of 16th International Conference on Pattern Recognition*, volume 4, pages 367–372, 2002.

[48] M. Zampato, R. Pistellato, and D. Maddalena. Visual motion estimation for tumbling satellite capture. In *Proceedings of British Machine Vision Conference*, pages 68.1–68.10, 1996.

[49] Sean Augenstein and Stephen M. Rock. Improved frame-to-frame pose tracking during vision-only SLAM/SFM with a tumbling target. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 3131–3138, 2011.

[50] Stephen Se, Piotr Jasiobedzki, and Richard Wildes. Stereo-vision-based 3D modeling of space structures. In *Proceedings of Sensors and Systems for Space Applications*, pages 65550E–65550E–13, 2007.

[51] C.P. Lu, G.D. Hager, and E. Mjolsness. Fast and globally convergent pose estimation from video images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(6):610–622, June 2000.

[52] Philip David, Daniel Dementhon, Ramani Duraiswami, and Hanan Samet. Softposit: Simultaneous pose and correspondence determination. *International Journal of Computer Vision*, 59(3):259–284, September 2004.

[53] J.C. Diaz and M. Abderrahim. Modified softposit algorithm for 3d visual tracking. In *Proceedings of IEEE International Symposium on Intelligent Signal Processing*, pages 1–6, October 2007.

[54] David G. Lowe. Fitting parameterized three-dimensional models to images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5):441–450, May 1991.

[55] Andrew I. Comport, Eric Marchand, and Franccois Chaumette. A real-time tracker for markerless augmented reality. In *Proceedinsg of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 36, 2003.

[56] Eric Marchand and Francois Chaumette. Virtual visual servoing: a framework for real-time augmented reality. In *Proceedings of European Association Conference on Computer Graphics*, volume 21, pages 289–298, 2002.

[57] Reza Safaee-Rad, Ivo Tchoukanov, Kenneth Carless Smith, and Bensiyon Benhabib. Three-dimensional location estimation of circular features for machine vision. *IEEE Transactions on Robotics and Automation*, 8(5):624–640, 1992.

[58] Christophe Doignon and Michel de Mathelin. A degenerate conic-based method for a direct fitting and 3-d pose of cylinders with a single perspective view. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 4220–4225, 2007.

[59] Chris Harris and Carl Stennet. Rapid a video rate object tracker. In *Proceedings of the British Machine Vision Conference*, pages 73–77, September 1990.

[60] Tom Drummond and Roberto Cipolla. Visual tracking and control using lie algebras. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2652–2659, 1999.

[61] Wolfgang Sepp. A direct method for real-time tracking in 3-d under variable illumination. In *Proceedings of 27th DAGM Symposium on Pattern Recognition*, pages 246–253, 2005.

[62] Amaury Dame and Eric Marchand. Accurate real-time tracking using mutual information. In *Proceedings of 9th IEEE International Symposium on Mixed and Augmented Reality*, pages 47–56, 2010.

[63] Guillaume Caron, Amaury Dame, and Éric Marchand. Direct model based visual tracking and pose estimation using mutual information. *Journal of Image, Vision and Computing*, 32(1):54–63, 2014.

[64] Luca Vacchetti, Vincent Lepetit, and Pascal Fua. Combining edge and texture information for real-time accurate 3d camera tracking. In *Proceedings of 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 48–56, 2004.

[65] Muriel Pressigout and Eric Marchand. Real-time 3d model-based tracking: Combining edge and texture information. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 2726–2731, 2006.

[66] C. Choi and H. I. Christensen. Robust 3D visual tracking using particle filtering on the special euclidean group: A combined approach of keypoint and edge features. *The International Journal of Robotics Research*, 31(4):498–519, March 2012.

[67] Michael Haag and Hans-Hellmut Nagel. Combination of edge element and optical flow estimates for 3D-model-based vehicle tracking in traffic image sequences. *International Journal of Computer Vision*, 35(3):295–319, 1999.

[68] Thomas Brox, Bodo Rosenhahn, Daniel Cremers, and Hans-Peter Seidel. High accuracy optical flow serves 3-d pose tracking: exploiting contour and flow based constraints. In *Proceedings of European conference on Computer Vision*, pages 98–111, 2006.

[69] Muriel Pressigout, Eric Marchand, and Etienne Memin. Hybrid tracking approach using optical flow and pose estimation. In *Proceedings of 15th IEEE International Conference on Image Processing*, pages 2720–2723, 2008.

[70] Edward Rosten and Tom Drummond. Fusing points and lines for high performance tracking. In *Proceedings of 10th IEEE International Conference on Computer Vision*, volume 2, pages 1508–1515, 2005.

[71] Karl Pauwels, Leonardo Rubio, Javier Diaz, and Eduardo Ros. Real-time model-based rigid object pose estimation and tracking combining dense and sparse visual cues. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 2347–2354, 2013.

[72] Ville Kyrki and Danica Kragic. Integration of model-based and model-free cues for visual object tracking in 3d. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, pages 1554–1560, 2005.

[73] Georg Klein and David W. Murray. Full-3D edge tracking with a particle filter. In *Proceedings of British Machine Vision Conference*, pages 1119–1128, 2006.

[74] Celine Teuliere, Eric Marchand, and Laurent Eck. Using multiple hypothesis in model-based tracking. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 4559–4565, 2010.

[75] Changhyun Choi and Henrik I. Christensen. Robust 3D visual tracking using particle filtering on the SE (3) group. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 4384–4390, 2011.

[76] Andrea Fossati and Pascal Fua. Linking pose and motion. In *Proceedings of European Conference on Computer Vision*, pages 200–213. 2008.

[77] Victor Adrian Prisacariu, Radu Timofte, Karel Zimmermann, Ian Reid, and Luc Van Gool. Integrating object detection with 3D tracking towards a better driver assistance system. In *Proceedings of 20th International Conference on Pattern Recognition*, pages 3344–3347, 2010.

[78] Victor A. Prisacariu and Ian D. Reid. Pwp3d: Real-time segmentation and tracking of 3d objects. *International Journal of computer vision*, 98(3):335–354, 2012.

[79] Christian Schmaltz, Bodo Rosenhahn, Thomas Brox, Daniel Cremers, Joachim Weickert, Lennart Wietzke, and Gerald Sommer. Region-based pose tracking. In *Proceedings of 3rd Iberian Conference on Pattern Recognition and Image Analysis*, pages 56–63, Girona, Spain, 2007.

[80] Samuel Dambreville, Romeil Sandhu, Anthony Yezzi, and Allen Tannenbaum. Robust 3d pose estimation and efficient 2d region-based segmentation from a 3d shape prior. In *Proceedings of European Conference on Computer Vision*, pages 169–182, 2008.

[81] Giorgio Panin, Alexander Ladikos, and Alois Knoll. An efficient and robust real-time contour tracking system. In *Proceedings of IEEE International Conference on Computer Vision Systems*, pages 44–44, 2006.

[82] Shulei Zhu, Dejan Pangercic, and Michael Beetz. Contracting curve density algorithm for applications in personal robotics. In *Proceedings of 11th IEEE-RAS International Conference on Humanoid Robots*, pages 171–178, 2011.

[83] P. Jasiobedzki, M. Greenspan, and G. Roth. Robust 3d vision for autonomous space robotic operation. In *Proceedings of 6th International Symposium on Artificial Intelligence, Robotics and Automation in Space*, June 2001.

[84] Fuyuto Terui. Model based visual relative motion estimation and control of a spacecraft utilizing computer graphics. In *Proceedings of 21st International Symposium on Space Flight dynamics*, Tolouse, France, 2009.

[85] Zhengyou Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of computer vision*, 13(2):119–152, 1994.

[86] Antoine Petit, Eric Marchand, and Keyvan Kanani. Tracking complex targets for space rendezvous and debris removal applications. In *Proceedings of IEEE International Conference on Intelligent Robots and Systems*, pages 4483–4488, 2012.

[87] C. Miravet, L. Pascual, E. Krouch, and J.M. Delcura. An image-based sensor system for autonmous rendez-vous with uncooperative satellites. In *Proceedings of ESA 7th International Conference on Guidance, Navigation and Control Systems*, June 2008.

[88] Wenfu Xu, Bin Liang, Bing Li, and Yangsheng Xu. A universal on-orbit servicing system used in the geostationary orbit. *Journal of Advances in Space Research*, 48(1):95–119, July 2011.

[89] Xuehai Gao, Bin Liang, and Wenfu Xu. Attitude determination of large non-cooperative spacecrafts in final approach. In *Proceedings of IEEE 11th International Conference on Control, Automation, Robotics and Vision*, pages 1571–1576, 2010.

[90] Alexande Cropp and P. Palmer. Pose estimation and relative orbit determination of a nearby target microsatellite using passive imagery. In *Proceedings of 5th Cranfield Conference on Dynamics and Control of Systems and Structures in Space*, pages 389–395, 2002.

[91] Simone Damico, Mathias Benn, and John Leif JOrgensen. Pose estimation of an uncooperative spacecraft from actual space imagery. In *Proceedings of 5th International Conference on Spacecraft Formation Flying Missions and Technologies*, Munich, Germany, 2013.

[92] Matthew D. Lichter and Steven Dubowsky. Estimation of state, shape, and inertial parameters of space objects from sequences of range images. In *Proceedings of SPIE on Photonics Technologies for Robotics, Automation, and Manufacturing*, pages 194–205, 2003.

[93] Winkler Wunsch, S. Winkler, P. Wunsch, and G. Hirzinger. A feature map approach to real-time 3-d object pose estimation from single 2-d perspective views. In *Proceedings of the German Association for Pattern Recognition*, pages 129–136, 1997.

[94] Frank Dellaert, Chuck Thorpe, and Sebastian Thrun. Super-resolved texture tracking of planar surface patches. In *In Proceedings of IEEE/RSJ International Conference on Intelligent Robotic Systems*, pages 197–203, 1998.

[95] Bruno Cernuschi-Frias, David B. Cooper, Yi ping Hung, and Peter N. Belhumeur. Toward a model-based bayesian theory for estimating and recognizing parameterized 3-d objects using two or more images taken from different positions. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 11(10):1028–1052, 1989.

[96] F. Jurie and Michel Dhome. A simple and efficient template matching algorithm. In *Proceedings of IEEE International Conference on Computer Vision*, volume 2, pages 544–549, 2001.

[97] J.M. Buenaposada and L.Baumela. Real-time tracking and estimation of plane pose. In *Proceedings of 16th International Conference on Pattern Recognition.*, volume 2, pages 697–700, 2002.

[98] Ali Shahrokni, Tom Drummond, and Pascal Fua. Texture boundary detection for real-time tracking. In *Proceedings of European Conference on Computer Vision*, pages 566–577, 2004.

[99] Dana Cobzas and Martin Jagersand. 3d ssd tracking from uncalibrated video. In *Proceedings of 1st International Workshop on Spatial coherence for visual motion analysis*, pages 25–37, 2006.

[100] Lucie Masson, Michel Dhome, and F. Jurie. Robust real time tracking of 3d objects. In *Proceedingss of International Conference on Pattern Recognition*, pages 252–255, 2004.

[101] Selim Benhimane and E. Malis. Real-time image-based tracking of planes using efficient second-order minimization. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 1, pages 943–948, September 2004.

[102] T. Moritani, S. Hiura, and K. Sato. Real-time object tracking without feature extraction. In *Proceedings of 18th International Conference on Pattern Recognition*, volume 1, pages 747–750, 2006.

[103] Wolfgang Sepp. *Visual Servoing of Textured Free Form Objects in 6 Degrees of Freedom.* PhD thesis, Technical University of Munich, 2008.

[104] V. Lepetit, P. Lagger, and P. Fua. Randomized trees for real-time keypoint recognition. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 775–781, June 2005.

[105] P. Wunsch and G. Hirzinger. Registration of cad-models to images by iterative inverse perspective matching. In *Proceedings of the 13th International Conference on Pattern Recognition*, volume 1, pages 78–83, 1996.

[106] Eric Marchand and Francois Chaumette. Features tracking for visual servoing purpose. *Journal of Robotics and Autonomous Systems*, 52(1):53–70, 2005.

[107] N.A. Ramey, J.J. Corso, W.W. Lau, D. Burschka, and G.D. Hager. Real-time 3d surface tracking and its applications. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition Workshop*, page 34, June 2004.

[108] P. Viola and W.M. I.I.I. Wells. Alignment by maximization of mutual information. In *Proceedings of 5th International Conference on Computer Vision*, pages 16–23, June 1995.

[109] James Paterson and Andrew Fitzgibbon. 3d head tracking using non-linear optimization. In *Proceedings of British Machine Vision Conference*, pages 609–618, 2003.

[110] Kim Junhwan, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. In *Proceedings of 9th IEEE International Conference on Computer Vision*, volume 2, pages 1033–1040, October 2003.

[111] D. Freedman and M.W. Turek. Illumination-invariant tracking via graph cuts. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 10–17, June 2005.

[112] Charles V. Stewart. Robust parameter estimation in computer vision. *Journal of Society for Industrial and Applied Mathematics (SIAM) Reviews*, 41(3):513–537, 1999.

[113] S. Romdhani and T. Vetter. Efficient, robust and accurate fitting of a 3d morphable model. In *Proceedings of 9th IEEE International Conference on Computer Vision*, volume 1, pages 59–66, October 2003.

[114] M. Pressigout and E. Marchand. Real time planar structure tracking for visual servoing: a contour and texture approach. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 251–256, August 2005.

[115] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the DARPA Image Understanding Workshop*, pages 674–679, 1981.

[116] V. Blanz, S. Romdhani, and T. Vetter. Face identification across different poses and illuminations with a 3d morphable model. In *Proceedings of 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 192–197, May 2002.

[117] Gregory D. Hager and P.N. Belhumeur. Real-time tracking of image regions with changes in geometry and illumination. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 403–410, June 1996.

[118] T.F. Cootes, K. Walker, and C.J. Taylor. View-based active appearance models. In *Proceedings of 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 227–232, 2000.

[119] Iain Matthews and Simon Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, November 2004.

[120] Peter N. Belhumeur and Gregory D. Hager. Tracking in 3d: Image variability decomposition for recovering object pose and illumination. *Journal of Pattern Analysis and Applications*, 2(1):82–91, 1999.

[121] Marco La Cascia, Stan Sclaroff, and Vassilis Athitsos. Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped

3d models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22 (4):322–336, April 2000.

[122] Jing Xiao, Simon Baker, Iain Matthews, and Takeo Kanade. Real-time combined 2d+3d active appearance models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 535– 542, June 2004.

[123] Jose Miguel Buenaposada, Enrique Munoz, and Luis Baumela. Efficient appearance-based tracking. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition Workshop*, volume 1, page 6, 2004.

[124] Dana Cobzas and Martin Jagersand. Tracking and predictive display for a remote operated robot using uncalibrated video. In *Proceedings of IEEE International conference on Robotics and Automation*, pages 1847–1852, April 2005.

[125] Enrique Munoz, Jose M. Buenaposada, and Luis Baumela. Efficient model-based 3d tracking of deformable objects. In *Proceeding of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 877–882, 2005.

[126] Jose M. Buenaposada, Enrique Munoz, and Luis Baumela. Efficient illumination independent appearance-based face tracking. *Journal of Image, Vision and Computing*, 27(5):560–578, April 2009.

[127] Soo-Chang Pei and Lin-Gwo Liou. Tracking a planar patch in three-dimensional space by affine transformation in monocular and binocular vision. *Journal of Pattern Recognition*, 26(1):23–31, 1993.

[128] Duan Suolin, Chen Lanping, Ma Zhenghua, and Yang Zhongyao. Moment invariant-based multi-target recognition and grasping for robot manipulator. In *Proceedings of 8th World Congress on Intelligent Control and Automation*, pages 1000–1005, July 2010.

[129] A. Yeremou Tamtsia, O. Tahri, Y. Mezouar, H. Djalo, and E. Tonye. New results in images moments-based visual servoing. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 5271–5276, May 2013.

[130] A.H.A. Hafez and C.V. Jawahar. Target model estimation using particle filters for visual servoing. In *Proceedings of 18th International Conference on Pattern Recognition*, volume 4, pages 651–654, 2006.

[131] Ezio Malis and François Chaumette. 2 1/2 d visual servoing with respect to unknown objects through a new estimation scheme of camera displacement. *International Journal of Computer Vision*, 37(1):79–97, June 2000.

[132] J. Stavnitzky and D. Capson. Multiple camera model-based 3-d visual servo. *IEEE Transactions on Robotics and Automation*, 16(6):732–739, December 2000.

[133] Robert Hanek and Michael Beetz. The contracting curve density algorithm: Fitting parametric curve models to images using local self-adapting separation criteria. *International Journal of Computer Vision*, 59(3):233–258, September 2004.

[134] Li Peihua, F. Chaumette, and O. Tahri. A shape tracking algorithm for visual servoing. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 2847–2852, April 2005.

[135] D. Kragic and V. Kyrki. Initialization and system modeling in 3-d pose tracking. In *Proceedings of 18th International Conference on Pattern Recognition*, volume 4, pages 643–646, 2006.

[136] Pascal Lagger, Mathieu Salzmann, Vincet Lepetit, and pascal Fua. 3d pose refinement from reflections. In *Proceedings of IEEE Conference on Computer vision ans Pattern Recognition*, pages 1–8, 2008.

[137] Aaron Netz and maragarita Osadchy. Using specular highlights as pose invariant features for 2d-3d pose estimation. In *Proceedings of IEEE Conference on Computer vision ans Pattern Recognition*, pages 721 –728, 2011.

[138] Ju Yong Chang, Ramesh Raskar, and Amit K. Agrawal. 3d pose estimation and segmentation using specular cues. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 1706–1713, 2009.

[139] Tianli Yu, Ning Xu, and Narendra Ahuja. Recovering shape and reflectance model of non-lambertian objects from multiple views. In *Proceedings of IEEE International Conference on Computer vision ans Pattern Recognition*, volume 2, pages II–226 – II–233, 2004.

[140] Michael Oren and Shree K. Nayar. A theory of specular surface geometry. 24(2): 105–124, 1995.

[141] Stefan Roth and Michael J Black. Specular flow and the recovery of surface structure. In *Proceedings of IEEE International Conference on Computer vision ans Pattern Recognition*, volume 2, pages 1869–1876, 2006.

[142] Yuriy Vasilyev, Yair Adato, Todd Zickler, and Ohad Ben-Shahar. Dense specular shape from multiple specular flows. In *Proceedings of IEEE International Conference on Computer vision ans Pattern Recognition*, pages 1–8, 2008.

[143] Yair Adato, Todd zickeler, and Ohad Benn-Shahr. Toward robust estimation of specular flow. In *Proceedings of British Machine Vision Conference*, pages 22.1–22.1, 2010.

[144] Michèle Gouiffès, Christophe Collewet, Christine Fernandez-Maloigne, and Alain Trémeau. Feature points tracking: Robustness to specular highlights and lighting changes. In *Proceedings of 9th European Conference on Computer Vision*, pages 82–93, 2006.

[145] R. Y. Tsai. An efficient and accurate camera calibration technique for 3d machine vision. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 364–374, 1986.

[146] Richard Hartley and Andrew Zisserman. Multiple view geometry in computer vision. *Cambridge University Press, Second Edition*, 2003.

[147] Klaus H. Strobl and G. Herzinger. optimal hand-eye calibration. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4647–4653, 2006.

[148] Edward Rosten, R. Porter, and Tom Drummond. Faster and better: A machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):105–119, January 2010.

[149] F. John Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[150] T. Lindeberg. Discrete derivative approximation with scale-space: A basis for low-level feature extraction. *International Journal of Mathematical Imaging and Vision*, 11(3):283–318, December 1993.

[151] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *Proceedings of IEEE International Conference on Computer Vision*, pages 2564–2571, 2011.

[152] Vincent Lepetit and Francesc Moreno-Noguer. Epnp: An accurate o(n) solution to the pnp problem. *International Journal of Computer Vision*, 81(2):155–166, 2008.

[153] Xiao shan Gao. Complete solution classification for the perspective-three-point problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25 (8):930–943, 2003.

[154] Luis Ferraz, Xavier Binefa, and Francesc Moreno-Noguer. Very fast solution to the pnp problem with algebraic outlier rejection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 501–508, June 2014.

[155] P.J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[156] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the icp algorithm. In *Proceedings of IEEE International Conference on 3-D Digital Imaging and Modeling*, pages 145–152, 2001.

[157] Antoine Petit, Eric Marchand, and Keyvan Kanani. Vision-based detection and tracking for space navigation in a rendezvous context. In *International Symposium on Artificial Intelligence, Robotics and Automation in Space, i-SAIRAS*, Turin, Italy, 2012.

[158] Toralf Boge and Heike Benninghoff. Rendezvous simulation for on-orbit servicing missions using advanced robotic technology. In *Proceedings of 19th International Symposium on Automatic Control in Space*, Wuerzburg, Germany, September 2013.

[159] Shahriar Negahdaripour. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(9):961–979, 1998.

[160] S Baker and I. Mathews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–225, 2004.

[161] Timo Zinsser, Christoph Grä ssl, and Heinrich Niemann. Effecient feature tracking for long video sequences. In *Proceedings of 26th DAGM Symposium on Pattern Recognition*, pages 326–333, 2008.

[162] Nassir W. Oumer and G. Panin. 3d point tracking and pose estimation of a space object using stereo images. In *Proceedings of 21st International Conference on Pattern Recognition*, pages 796–800, November 2012.

[163] Nassir W. Oumer and G. Panin. Tracking and pose estimation of a non-cooperative satellite for on-orbit servicing. In *Proceedings of International Symposium on Artificial Intelligence, Robotics and Automation in Space*, Turin, Italy, September 2012.

[164] Allen M. WaxMan and James H. Duncan. Binocular image flows: Steps toward stereo-motion fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):715–729, 1986.

[165] Shahriar Negahdaripour. Motion recovery from images sequences using only first order optical flow information. *International Journal of Computer Vision*, pages 163–184, 1992.

[166] Nassir.W. Oumer and Giorgio Panin. Camera-based tracking for rendezvous and proximity operation of a satellite. In *Proceedings of Advances in Aerospace Guidance, Navigation and Control*, pages 625–638, Toulouse, France, April 2015.

[167] Nassir W. Oumer, Giorgio Panin, Quirin Mühlbauer, and Anastasia Tseneklidou. Vision-based localization for on-orbit servicing of a partially cooperative satellite. *Journal of Acta Astronautica*, 117:679–698, 2015.

[168] J. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7(4):308–313, 1965.

[169] David G. Lowe. Fitting parametrized three-dimensional models to images. *IEEE Transaction Pattern Analysis and Machine Intelligence*, 13:441–450, 1991.

[170] I. Ladrón De Guevara, J. Mu noz, O. D. Cózar, and E. B. Blázquez. Robust fitting of circle arcs. *Journal of Mathematical Imaging and Vision*, 40(2):147–161, June 2011.

[171] Andrew I. Comport, Éric Marchand, Muriel Pressigout, and François Chaumette. Real-time markerless tracking for augmented reality: The virtual visual servoing framework. *IEEE Transaction on Visualization and Computer Graphics*, 12(4): 615–628, 2006.

[172] H. Jin and S. Soatto. Structure from motion causally integrated over time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):523–535, 2002.

[173] R. Lampariello, N. W. Oumer, J. Artigas, W. Rackl, G. Panin, R. Purschke, J. Harder, U. Walter, J. Frickel, I. Masic, K. Ravandoor, J. Scharnagl, K. Schilling, K. Landzettel, and G. Hirzinger. Forrost: Advances in on-orbit robotic technologies. In *Proceedings of IEEE Aerospace Conference*, pages 1–20, March 2015.

[174] Marcus Ulrich, Christian Wiedemann, and Carsten Steger. Cad-based recognition of 3d objects in monocular images. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 1191–1198, Kobe, 2009.

[175] Antoine Petit, Eric Marchand, Rafiq Sekkal, and Keyvan Kanani. 3d object pose detection using foreground/background segmentation. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 1858 –1865, Seattle, WA, 2015.

[176] Georg Klein. *Visual Tracking for Augmented Reality*. PhD thesis, University of Cambridge, 2006.

[177] Christopher M. Cyr and Benjamin B. Kimia. 3d object recognition using shape similiarity-based aspect graph. In *Proceedings of 8th IEEE International Conference on Computer Vision*, volume 1, pages 254 –261, Vancouver, BC, 2001.

[178] Christian Reinbacher, Matthias Rüther, and Horst Bischof. Pose estimation of known objects by efficient silhouette matching. In *Proceedings of International Conference on Pattern Recognition*, pages 1080–1083, 2010.

[179] S. Hinterstoisser, S. Benhimane, and N. Navab. N3m: Natural 3d markers for real-time object detection and pose estimation. In *Proceedings of the IEEE 11th International Conference on Computer Vision*, pages 1–7, Rio de Janeiro, 2007.

[180] David. G. Lowe. Three-dimensional object recognition from single two-dimensional images. *Journal of Artificial Intelligence*, 31(3):335–395, 1987.

[181] M. S. Costa and L. G. Shapiro. 3d object recognition and pose with relational indexing. *Journal of Computer Vision and Image Understanding*, 79(3):364–407, 2000.

[182] P. David and D. DeMenthon. Simultaneous pose and correspondence determination using line features. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 424–431, 2003.

[183] P. David and D. DeMenthon. Object recognition in high clutter images using line features. In *Proceedings of 10th IEEE International Conference on Computer Vision*, pages 1581–1588, 2005.

[184] Vincent Lepetit. Keypoint recognition using randomized trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1465–1479, 2006.

[185] David Nister and Henrik Stewenius. Scalable recognition with a vocabulary tree. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 2161–2168, 2006.

[186] Heung-Yeung Shum, Shing-Chow Chan, and Sing Bing Kang. Image-based rendering. *Springer*, 2007.

[187] R. Arandjelovic and A. Zisserman. Three things everyone should know to improve object retrieval. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 2911–2918, 2012.

[188] Tinne Tuytelaars and Krystian Mikolajczyk. Local invariant feature detectors: A survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2007.