CNN-Based Pose Estimation System for Close-Proximity Operations Around Uncooperative Spacecraft

L. Pasqualetto Cassinis *, R. Fonod † and E. Gill ‡
Delft University of Technology, Kluyverveg 1, 2629 HS, Delft, The Netherlands

I. Ahrns §
Airbus DS GmbH, Airbusallee 1, 28199, Bremen, Germany

J. Gil Fernandez¶
ESTEC, Keplerlaan 1, 2201 AZ, Noordwijk, The Netherlands

This paper introduces a novel framework which combines a Convolutional Neural Network (CNN) for feature detection with a Covariant Efficient Procrustes Perspective-n-Points (CEPPnP) solver and an Extended Kalman Filter (EKF) to enable robust monocular pose estimation for close-proximity operations around an uncooperative spacecraft. The relative pose estimation of an inactive spacecraft by an active servicer spacecraft is a critical task in the design of current and planned space missions, due to its relevance for close-proximity operations, such as In-Orbit Servicing and Active Debris Removal. The main contribution of this work stands in deriving statistical information from the Image Processing step, by associating a covariance matrix to the heatmaps returned by the CNN for each detected feature. This information is included in the CEPPnP to improve the accuracy of the pose estimation step during filter initialization. The derived measurement covariance matrix is used in a tightly-coupled EKF to allow an enhanced representation of the measurements error from the feature detection step. This increases the filter robustness in case of inaccurate CNN detections. The proposed method is capable of returning reliable estimates of the relative pose as well as of the relative translational and rotational velocities, under adverse illumination conditions and partial occultation of the target. Synthetic 2D images of the European Space Agency's Envisat spacecraft are used to generate datasets for training, validation and testing of the CNN. Likewise, the images are used to recreate representative close-proximity scenarios for the validation of the proposed method.

I. Introduction

Nowadays, key Earth-based applications such as remote sensing, navigation, and telecommunication, rely on satellite technology on a daily basis. To ensure a high reliability of these services, the safety and operations of satellites in orbit has to be guaranteed. In this context, advancements in the field of Guidance, Navigation, and Control (GNC) were made in the past years to cope with the challenges involved in In-Orbit Servicing (IOS) and Active Debris Removal (ADR) missions [1, 2]. For such scenarios, the estimation of the relative position and attitude (pose) of an uncooperative spacecraft by an active servicer spacecraft represents a critical task. Pose estimation systems based solely on a monocular camera are recently becoming an attractive alternative to systems based on active sensors or stereo cameras, due to their reduced mass, power consumption and system complexity [3]. However, given the low Signal-To-Noise Ratio (SNR) and the high contrast which characterize space images, a significant effort is still required to comply with most of the demanding requirements for a robust and accurate monocular-based navigation system. Interested readers are referred to Ref. [4] for a recent overview of the current trends in monocular-based pose estimation systems. Notably, the pose estimation problem is in this case complicated by the fact that the target satellite is uncooperative, namely retained as not functional and/or not able to aid the relative navigation. Above all, the navigation system cannot rely

*PhD Candidate, Department of Space Engineering, *L.PasqualettoCassinis@tudelft.nl*
†Assistant Professor, Department of Space Engineering, *R.Fonod@tudelft.nl*
‡Professor, Department of Space Engineering, *E.K.A.Gill@tudelft.nl*
§Engineer, Space Robotics Projects, *ingo.ahrns@airbus.com*
¶Engineer, TEC-ECN, *Jesus.Gil.Fernandez@esa.int*

on known visual markers, as they are typically not installed on an uncooperative target. Since the extraction of visual features is an essential step in the pose estimation process, advanced Image Processing (IP) techniques are required to extract keypoints (or interest points), corners, and/or edges on the target body. In model-based methods, the detected features are then matched with pre-defined features on an offline wireframe 3D model of the target to solve for the relative pose. In other words, a reliable detection of key features under adverse orbital conditions is highly desirable to guarantee safe operations around an uncooperative spacecraft. Moreover, it would be beneficial from a different standpoint to obtain a model of feature detection uncertainties. This would provide the navigation system with additional statistical information about the measurements, which could in turn improve the robustness of the entire estimation process.

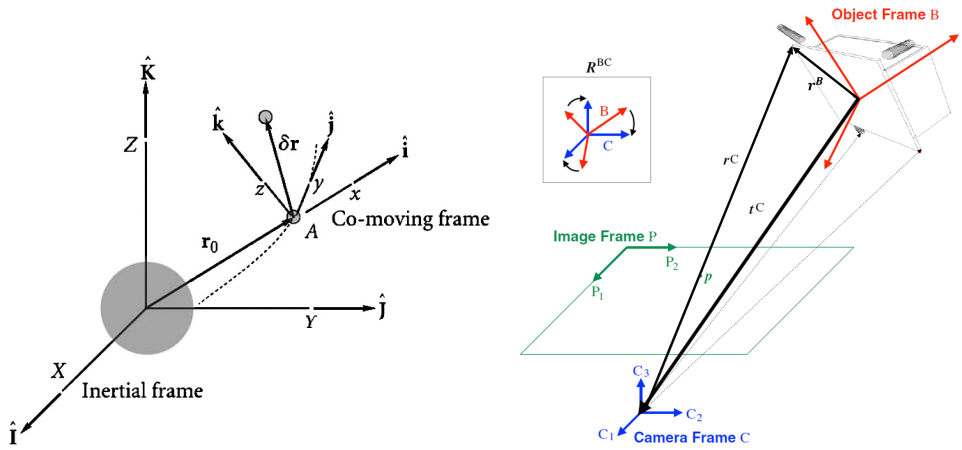
Unfortunately, standard pose estimation solvers such as the Efficient Perspective-n-Point (EPnP) [5], the Efficient Procrustes Perspective-n-Point (EPPnP) [6], or the multi-dimensional Newton Raphson Method (NRM) [7] do not have the capability to include features uncertainties. Only recently, a Covariant EPPnP (CEPPnP) solver was introduced to exploit statistical information by including feature covariances in the pose estimation [8]. The authors proposed a method for computing the covariance which takes different camera poses to create a fictitious distribution around each detected keypoint. Furthermore, other authors [9] proposed an improved pose estimation method based on projection vector in which the covariance is associated to the image gradient magnitude and direction at each feature location. However, in both methods the derivation of features covariance matrices cannot be directly related to the actual detection uncertainty. In this context, Convolutional Neural Networks (CNN) could be exploited to return relevant statistical information about the detection step. This could in turn provide a reliable representation of the detection uncertainty.

The implementation of CNNs for monocular pose estimation in space has already become an attractive solution in recent years [10–12], also thanks to the creation of the Spacecraft PosE Estimation Dataset (SPEED) [11], a database of highly representative synthetic images of PRISMA's TANGO spacecraft made publicly available by Stanford's Space Rendezvous Laboratory (SLAB) applicable to train and test different network architectures. One of the main advantages of CNNs over standard feature-based algorithms for relative pose estimation [3, 13, 14] is an increase in the robustness under adverse illumination condition, as well as a reduction in the computational complexity. Since the pose accuracies of the first adopted CNNs proved to be lower than the accuracies returned by common pose estimation solvers, especially in the estimation of the relative attitude [10], recent efforts investigated the capability of CNNs to perform keypoint localization prior to the actual pose estimation [15–18]. The output of these networks is a set of so called *heatmaps* around pre-trained features. The coordinates of the heatmap's peak intensity characterize the predicted feature location, with the intensity and the shape indicating the confidence of locating the corresponding keypoint at this position [15]. Additionally, due to the fact that the trainable features can be selected offline prior to the training, the matching of the extracted feature points with the features of the wireframe model can be performed without the need of a large search space for the image-model correspondences, which usually characterizes most of the edges/corners-based methods [19].

To the best of the authors' knowledge, the reviewed implementations of CNNs feed solely the heatmap's peak location into the pose estimation solver, despite multiple information could be extracted from the detected heatmaps. Only in Ref. [15], the pose estimation is solved by assigning weights to each feature based on their heatmap's peak intensities, in order to penalize inaccurate detections. Yet, there is another aspect related to the heatmaps which has not been considered. It is in fact hardly acknowledged how the overall shape of the detected heatmaps returned by CNN can be translated into a statistical distribution around the peak, allowing reliable feature covariances and, in turn, a robust navigation performance. Despite the general claim that visual-based navigation filters can generally work with an a-priori, overestimated, constant measurement covariance matrix, an accurate representation of the measurements uncertainty can in fact be beneficial also for the estimation of the relative state vector, which would include the relative pose as well as the relative translational and rotational velocities.

In this framework, the main contribution of this paper is to combine a CNN-based feature detector with a CEPPnP solver and an Extended Kalman Filter (EKF). The filter accounts for representative feature covariances, computed directly from the statistical distribution derived from the CNN heatmaps, and is initialized by the CEPPnP at time t_0. Specifically, the novelty of this work stands in linking the current research on CNN-based feature detection, covariant-based PnP solvers, and navigation filters for relative pose estimation. The work is driven by two main objectives:

1) To improve filter initialization by incorporating heatmaps-derived covariance matrices in the CEPPnP
2) To guarantee a robust estimation by including a representative measurement covariance matrix in the EKF.



(a) Co-moving LVLH frame [20]

(b) PnP problem (Figure adapted from [3])

Fig. 1 Representation of the relative motion framework (left), and schematic of the pose estimation problem using a monocular image (right).

The paper is organized as follows. The overall pose estimation framework is illustrated in Section II. Section III introduces the proposed CNN architecture together with the adopted training, validation, and testing datasets. In Section IV, special focus is given to the derivation of covariance matrices from the CNN heatmaps, whereas Section V describes the CEPPnP solver. Besides, Section VI provides a description of the adopted EKF. The simulation environment is presented in Section VII together with the simulation results. Finally, Section VIII provides the main conclusions and recommendations.

II. Pose Estimation Framework

This work considers a servicer spacecraft flying in relative motion around a target spacecraft located in a Low Earth Orbit (LEO), with the relative motion being described in a Local Vertical Local Horizontal (LVLH) reference frame co-moving with the servicer (Figure 1a). Furthermore, it is assumed that the servicer is equipped with a single monocular camera. The relative attitude of the target with respect to the servicer can then be defined as the rotation of the target body-fixed frame B with respect to the servicer camera frame C, where these frames are tied to each spacecraft's body. The distance between these two frames defines their relative position. Together, these two quantities characterize the relative pose. This information can then be transferred from the camera frame to the servicer's center of mass by accounting for the relative pose of the camera with respect to the LVLH frame.

From a high-level perspective, a model-based monocular pose estimation system receives as input a 2D image and matches it with an existing wireframe 3D model of the target spacecraft to estimate the pose of such target with respect to the servicer camera. Referring to Figure 1b, the pose estimation problem consists in determining the position of the target's centre of mass $\mathbf{t}^C$ and its orientation with respect to the camera frame C, represented by the rotation matrix $\mathbf{R}_B^C$. The Perspective-n-Points (PnP) equations,

$$\mathbf{r}^C = \begin{pmatrix} x^C & y^C & z^C \end{pmatrix}^T = \mathbf{R}_B^C \mathbf{r}^B + \mathbf{t}^C \tag{1}$$

$$\mathbf{p} = (u_i, v_i) = \left(\frac{x^C}{z^C} f_x + C_x, \frac{y^C}{z^C} f_y + C_y \right), \tag{2}$$

relate the unknown pose with the corresponding point $\mathbf{p}$ in the image plane. Here, $\mathbf{r}^B$ is a point in the 3D model, expressed in the body-frame coordinate system B, whereas f_x and f_y denote the focal lengths of the camera and (C_x, C_y) is the principal point of the image.

From these equations, it can already be seen that an important aspect of estimating the pose resides in the capability of the IP system to extract features $\boldsymbol{p}$ from a 2D image of the target spacecraft, which in turn need to be matched with

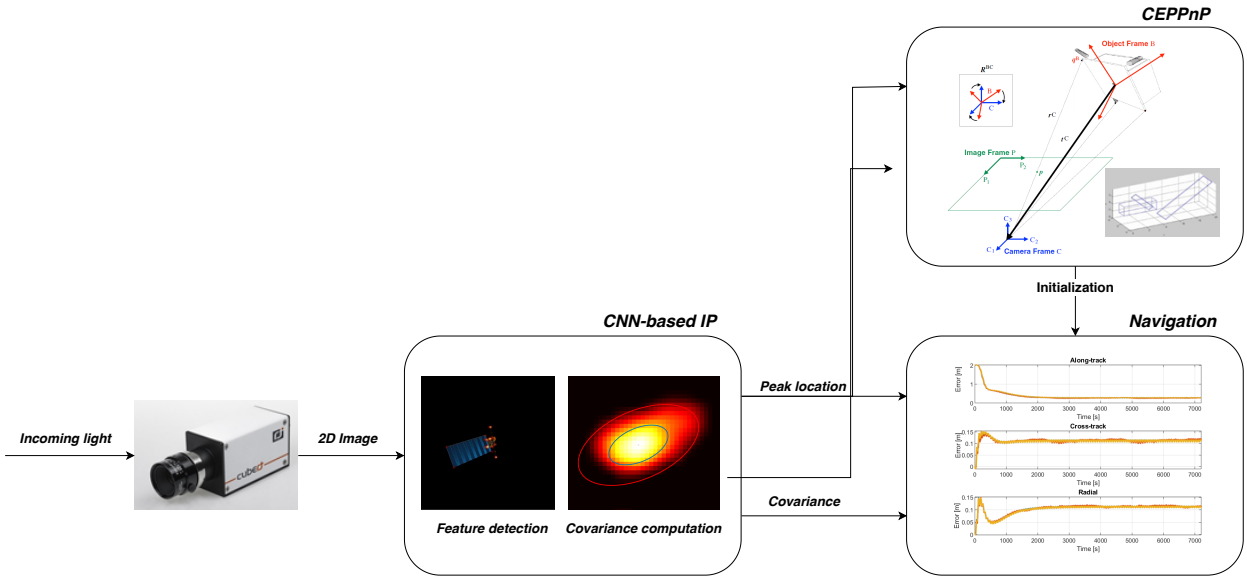


Fig. 2 Functional flow of the proposed pose estimation architecture.

pre-selected features $\boldsymbol{r}^B$ in the wireframe 3D model. Notably, such wireframe model of the target needs to be made available prior to the estimation. Furthermore, it can be expected that the time variation of the relative pose plays a crucial role while navigating around the target spacecraft, e.g. if rotational synchronization with the target spacecraft is required in the final approach phase. As such, it is clear that the estimation of both the relative translational and angular velocities is an essential step beside the pose estimation itself.

The proposed architecture combines the above key ingredients in three main stages, which are shown in Figure 2 and described in more detail in the following sections. In the CNN-based IP block, a CNN is used to extract features from a 2D image of the target spacecraft. Statistical information is derived by computing a covariance matrix for each features using the information included in the output heatmaps. In the Navigation block, both the peak locations and the covariances are fed into an EKF, which estimates the relative pose as well as the relative translational and rotational velocities. The filter is initialized by the CEPPnP block, which takes peak location and covariance matrix of each feature as input and outputs the initial relative pose by solving the PnP problem in Eqn.1-2. Thanks to the availability of a covariance matrix of the detected features, this architecture can guarantee a more accurate representation of feature uncertainties, especially in case of inaccurate detection of the CNN due to adverse illumination conditions and/or unfavourable relative geometries between servicer and target. Together with the CEPPnP initialization, this aspect can return a robust and accurate estimation of the relative pose and velocities and assure a safe approach of the target spacecraft.

III. Convolutional Neural Network

CNNs are currently emerging as a promising features extraction method, mostly due to the capability of their convolutional layers to extract high-level features of objects with improved robustness against image noise and illumination conditions. In order to optimize CNNs for the features extraction process, a stacked hourglass architecture has been proposed in Ref. [15, 16], and other architectures such as the U-net [21], or variants of the hourglass, were tested in recent years. Compared to the network proposed in Ref. [15], the adopted architecture is composed of only one encoder/decoder block, constituting a single hourglass module. The encoder includes six blocks, each including a convolutional layer, a batch normalization module and max pooling layer, whereas the six decoder blocks accomodate an up-sampling block in spite of max pooling. Each convolutional layer is formed by a fixed number of filter kernels of size 3×3. In the current analysis, 128 kernels are considered per convolutional layer. Figure 3 shows the high-level architecture of the network layers, together with the corresponding input and output.

As already mentioned, the output of the network is a set of heatmaps around the selected features. Ideally, the

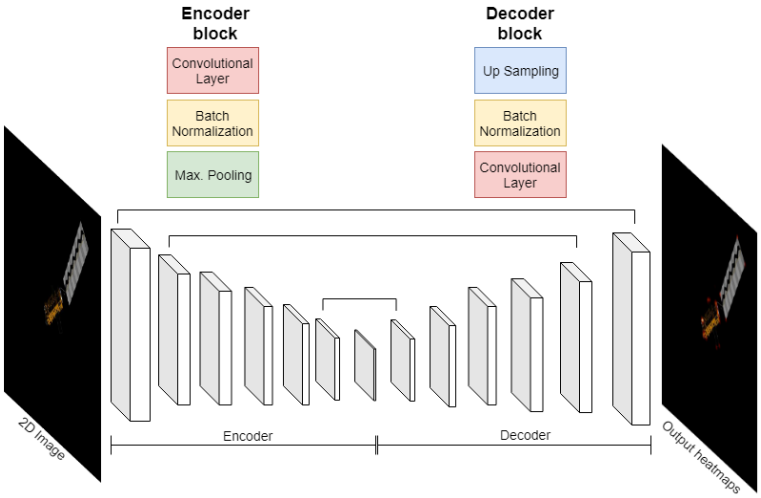


Fig. 3 Overview of the single hourglass architecture. Downsampling is performed in the *encoder* stage, in which the image size is decrease after each block, whereas upsampling occurs in the *decoder* stage. The output of the network consists of heatmap responses, and is used for keypoints localization.

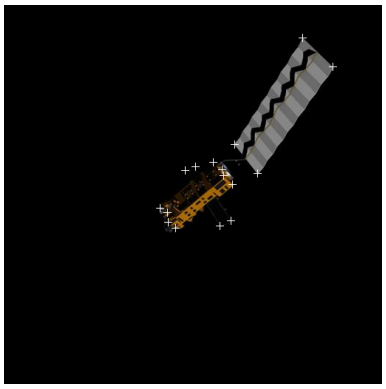


Fig. 4 Illustration of the selected features for a given Envisat pose.

heatmap's peak intensity associated to a wrong detection should be relatively small compared to the correctly detected features, highlighting that the network is not confident about that particular wrongly-detected feature. At the same time, the heatmap's amplitude should provide an additional insight into the confidence level of each detection, a large amplitude being related to large uncertainty about the detection. The network is trained with the x- and y- image coordinates of the feature points, computed offline based on the intrinsic camera parameters as well as on the feature coordinates in the target body frame, which were extracted from the wireframe 3D model prior to the training. During training, the network is optimized to locate 16 features of the Envisat spacecraft, consisting of the corners of the main body, the Synthetic-Aperture Radar (SAR) antenna, and the solar panel, respectively. Figure 4 illustrates the selected features for a specific target pose.

A. Training, Validation and Test

For the training, validation, and test datasets, synthetic images of the Envisat spacecraft were rendered in the Cinema 4D© software. Table 1 lists the main camera parameters adopted. A constant Sun elevation and azimuth angles of 30 degrees were chosen in order to recreate favourable as well as adverse illumination conditions. Relative distances between camera and target were chosen in the interval 90 m - 180 m. Relative attitudes were generated by discretizing the yaw, pitch, and roll angles of the target with respect to the camera by 10 degrees each. Together, these two choices were made in order to recreate several relative geometries between the servicer and the target. The resulting database was then shuffled to randomize the images, and was ultimately split into training (40,000 images), validation (4,000

Table 1 Parameters of the camera used to generate the synthetic images in Cinema 4D©.

Parameter	Value	Unit
Image resolution	512×512	pixels
Focal length	$3.9 \cdot 10^{-3}$	m
Pixel size	$1.1 \cdot 10^{-5}$	m

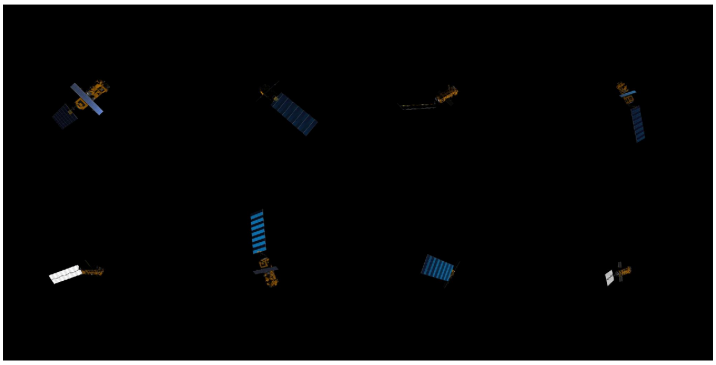


Fig. 5 A montage of eight synthetic images selected from the training set.

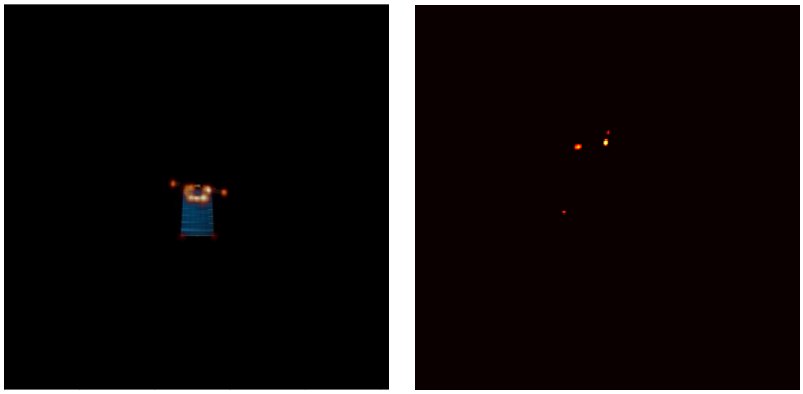
images), and test (~10,000 images) datasets. Figure 5 shows some of the images included in the training dataset. During training, the validation dataset is used beside the training one to compute the validation losses and avoid overfitting. In this work, a learning rate of 0.1 is selected together with a batch size of 20 images and a total number of 20 epochs. Finally, the network performance after training is assessed with the test dataset.

Preliminary results on the network performance were already reported in Ref. [22]. Above all, one key advantage of relying on CNNs for feature detection was found in the capability of learning the relative position between features under a variety of relative poses present in the training. As a result, both features which are not visible due to adverse illumination and features occulted by other parts of the target can be detected. Besides, a challenge was identified in the specific selection of the trainable features. Since the features selected in this work represent highly symmetrical points of the Envisat spacecraft, such as corners of the solar panel, SAR antenna or main body, the network is in some scenarios unable to distinguish between similar features, and returns multiple heatmaps for a single feature output. Figure 6 illustrates these findings.

Although improving on the CNN architecture is expected to cope with the above challenges and return more accurate heatmaps, it is also believed that these scenarios can be properly handled, if a large uncertainty can be associated to a wrong and/or inaccurate detection. This task can be performed by deriving a covariance matrix for each detected feature, in order to represent its detection uncertainty. Above all, this can prevent the pose solver and the EKF from trusting wrong detections, by relying more on other accurate features.

IV. Covariance Computation

In order to derive a covariance matrix associated to each feature from the heatmaps detected by the CNN, the first step is to obtain a statistical population around the heatmap's peak. This is done by thresholding each heatmap image so that only the x- and y- location of heatmap's pixels are extracted. Secondly, each pixel within the population is given a normalized weight w_i based on the gray intensity at its location. This is done in order to give more weight to pixels



(a) Occulted spacecraft body (b) Heatmap of a single solar panel's corner

Fig. 6 Robustness and challenges of feature detection with the proposed CNN. On the left-hand side, the network has been trained to recognize the pattern of the features, and can correctly locate the body features which are not visible, i.e. parts occulted by the solar panel and corners of the SAR antenna. Conversely, the right-hand side shows the detection of multiple heatmaps for a single corner of the solar panel. As can be seen, the network can have difficulties in distinguishing similar features, such as the corners of the solar panel.

which are particularly bright and close to the peak, and less weight to pixels which are very faint and far from the peak. Finally, the obtained statistical population of each feature is used to compute the weighted covariance between x, y and consequently the covariance matrix $\boldsymbol{C}_i$:

$$\boldsymbol{C}_i = \begin{pmatrix} \mathrm{cov}(x,x) & \mathrm{cov}(x,y) \\ \mathrm{cov}(y,x) & \mathrm{cov}(y,y) \end{pmatrix} \tag{3}$$

where

$$\mathrm{cov}(x,y) = \sum_{i=1}^{n} w_i (x_i - p_x) \cdot (y_i - p_y). \tag{4}$$

In this work, the mean is replaced by the peak location $\boldsymbol{p} = (p_x, p_y)$ in order to represent a distribution around the peak of the detected feature, rather than around the heatmap's mean. This is particularly relevant when the heatmaps are asymmetric and their mean does not coincide with their peak.

Figure 7 shows the overall flow to obtain the covariance matrix for three different heatmap shapes. The ellipse associated to each features covariance is obtained by computing the eigenvalues $\lambda_{x,y}$ of the covariance matrix,

$$\left(\frac{x}{\lambda_x}\right)^2 + \left(\frac{y}{\lambda_y}\right)^2 = s \tag{5}$$

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

V. Pose Estimation

The CEPPnP method proposed in Ref. [8] was selected to estimate the relative pose from the detected features as well as from their covariance matrices. The first step of this method is to rewrite the PnP problem in Eqn.1-2 as a function of a 12-dimensional vector $\boldsymbol{y}$ containing the control point coordinates in the camera reference system:

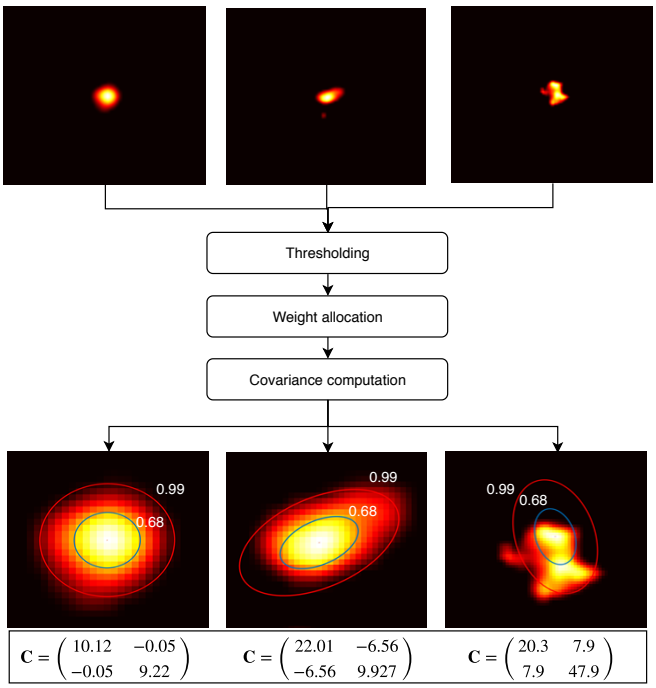


Fig. 7 Schematic of the procedure followed to derive covariance matrices from CNN heatmaps. The displayed ellipses are derived from the computed covariances by assuming the confidence intervals $1\sigma = 0.68$ and $3\sigma = 0.99$.

$$\boldsymbol{M}\boldsymbol{y} = \boldsymbol{0} \tag{6}$$

where $\boldsymbol{M}$ is a $2n\times12$ known matrix. This is the fundamental equation in the EPnP problem [5]. The likelihood of each observed feature location $\boldsymbol{u}_i$ is then represented as

$$P(\boldsymbol{u}_i) = k \cdot e^{-\frac{1}{2}\Delta\boldsymbol{u}_i^T \boldsymbol{C}_{\boldsymbol{u}_i}^{-1} \Delta\boldsymbol{u}_i} \tag{7}$$

where $\Delta\boldsymbol{u}_i$ is a small, independent and unbiased noise with expectation $E[\Delta\boldsymbol{u}_i] = 0$ and covariance $E[\Delta\boldsymbol{u}_i \Delta\boldsymbol{u}_i^T] = \sigma^2 \boldsymbol{C}_{\boldsymbol{u}_i}$ and k is a normalization constant. Here, σ^2 represents the global uncertainty in the image, whereas $\boldsymbol{C}_{\boldsymbol{u}_i}$ is the 2x2 uncertainty covariance matrix of each detected feature, computed from the CNN heatmaps. After some calculations, the EPnP formulation can be rewritten as

$$(\boldsymbol{N} - \boldsymbol{L})\boldsymbol{y} = \lambda \boldsymbol{y}. \tag{8}$$

This is an eigenvalue problem in which both $\boldsymbol{N}$ and $\boldsymbol{L}$ matrices are a function of $\boldsymbol{y}$ and $\boldsymbol{C}_{\boldsymbol{u}_i}$. The problem is solved iteratively by means of the closed-loop EPPnP solution for the four control points, assuming no feature uncertainty. Once $\boldsymbol{y}$ is estimated, the relative pose is computed by solving the generalized Orthogonal Procrustes problem used in the EPPnP [6].

VI. Navigation Filter

From a high level perspective, two different navigation architectures are normally exploited in the framework of relative pose estimation. A *tightly-coupled* architecture, where the extracted features are directly processed by the navigation filter as measurements, and a *loosely-coupled* architecture, in which the relative pose is computed prior to the navigation filter to derive pseudomeasurements from the target features [23]. Usually, a loosely-coupled approach is preferred for an uncooperative tumbling target, due to the fact that the fast relative dynamics could jeopardize feature tracking and return highly-variable measurements to the filter. However, one shortcoming of this approach is that it is generally hard to obtain a representative covariance matrix for the pseudomeasurements. This can be quite

challenging when filter robustness is demanded. Remarkably, the adoption of a CNN in the feature detection step can overcome the challenges in feature tracking by guaranteeing the detection of a constant, pre-defined set of features. At the same time, the CNN heatmaps can be used to derive a measurements covariance matrix and improve filter robustness.

Following this line of reasoning, a thighly-coupled EKF was chosen in this work. The state vector is composed of the relative pose between the servicer and the target, as well as the relative translational and rotational velocities. Under the assumption that the camera frame onboard the servicer is co-moving with the LVLH frame, with the camera boresight aligned with the along-track direction, this translates into

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{t}^C \\ \boldsymbol{v} \\ \boldsymbol{q} \\ \boldsymbol{w} \end{pmatrix} \tag{9}$$

where the unit quaternion $\boldsymbol{q}$ replaces the direction cosine matrix $\boldsymbol{R}_B^C$ to represent the coordinate transformation. The state propagation step decouples the translational dynamics from the rotational dynamics. The perturbation-free, closed-form solution of the well known Clohessy-Wiltshire linear equations was chosen to describe the relative motion between the servicer and the target:

$$\begin{pmatrix} \boldsymbol{t}^C \\ \boldsymbol{v} \end{pmatrix}_{k+1} = \boldsymbol{\Phi}_{CW} \begin{pmatrix} \boldsymbol{t}^C \\ \boldsymbol{v} \end{pmatrix}_k \tag{10}$$

where

$$\boldsymbol{\Phi}_{CW} = \begin{pmatrix} 1 & 0 & 6(\Delta\theta_s - \sin\Delta\theta_s) & 1/\omega_s(4\sin\Delta\theta_s - 3\Delta\theta_s) & 0 & 2/\omega_s(1-\cos\Delta\theta_s) \\ 0 & \cos\Delta\theta_s & 0 & 0 & 1/\omega_s \sin\Delta\theta_s & 0 \\ 0 & 0 & 4 - 3\cos\Delta\theta_s & 2/\omega_s(\cos\Delta\theta_s - 1) & 0 & \sin\Delta\theta_s/\omega_s \\ 0 & 0 & 6\omega_s(1-\cos\Delta\theta_s) & 4\cos\Delta\theta_s - 3 & 0 & 2\sin\Delta\theta_s \\ 0 & \omega_s \sin\Delta\theta_s & 0 & 0 & \cos\Delta\theta_s & 0 \\ 0 & 0 & 3\omega_s \sin\Delta\theta_s & -2\sin\Delta\theta_s & 0 & \cos\Delta\theta_s \end{pmatrix}. \tag{11}$$

Here, ω_s and $\Delta\theta_s$ represent the servicer's argument of perigee and true anomaly variation for a time step Δt, respectively. The relative quaternions and the rotational velocity are also propagated assuming a perturbation-free dynamics with constant angular velocity:

$$\boldsymbol{q}_{k+1} = \boldsymbol{q}_k \otimes \boldsymbol{q}(\boldsymbol{w}_k \Delta t) \tag{12}$$

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k \tag{13}$$

where the term $\boldsymbol{q}(\boldsymbol{w}_k \Delta t)$ represents the unit quaternion equivalent to a rotation of $\boldsymbol{w}_k \Delta t$. The propagation of the state covariance matrix $\boldsymbol{P}$ is done by accounting for a linearized version of Eqn.12. The reader is referred to Ref. [24] for a detailed description of this derivation.

In the measurements update step, the measurements vector is made of the x- and y- coordinates of the $n = 16$ detected features in the image plane,

$$\boldsymbol{z} = (x_1, y_1 \quad \dots \quad x_n, y_n)^T. \tag{14}$$

As anticipated, the measurement covariance matrix $\boldsymbol{R}$ is a time-varying block diagonal matrix constructed with the heatmaps-derived covariances $\boldsymbol{C}_i$ in Eqn.3,

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{C}_1 & & \\ & \ddots & \\ & & \boldsymbol{C}_n \end{pmatrix}. \tag{15}$$

Notice that $\boldsymbol{C}_i$ can differ for each feature in a given frame as well as vary over time.

The observation model $\boldsymbol{h}$ is a $2n\times1$ column vector representing the x- and y- coordinates of the 16 features detected by the CNN. Referring to Eqn.1-2, this translates into the following equations for each detected point $\boldsymbol{p}_i$:

$$\boldsymbol{h}_i = \left(\frac{x_i^C}{z_i^C} f_x + C_x, \frac{y_i^C}{z_i^C} f_y + C_y \right)^T \tag{16}$$

$$\mathbf{r}^C = \boldsymbol{q} \otimes \boldsymbol{r}_i^B \otimes \boldsymbol{q} + \mathbf{t}^C \tag{17}$$

where $\otimes$ denotes the quaternion product. As a result, the Jacobian $\boldsymbol{H}$ of the observation model with respect of the state vector is a $2n\times13$ matrix computed as

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_{\boldsymbol{t}^C,i} & \mathbf{0}_{2n\times3} & \boldsymbol{H}_{\boldsymbol{q},i} & \mathbf{0}_{2n\times3} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{H}_{\boldsymbol{t}^C,n} & \mathbf{0}_{2n\times3} & \boldsymbol{H}_{\boldsymbol{q},n} & \mathbf{0}_{2n\times3} \end{pmatrix} \tag{18}$$

where

$$\boldsymbol{H}_{r,i} = \boldsymbol{H}_i^{\text{int}} \cdot \boldsymbol{H}_{\boldsymbol{t}^C,i}^{\text{ext}} \tag{19}$$

$$\boldsymbol{H}_{\boldsymbol{q},i} = \boldsymbol{H}_i^{\text{int}} \cdot \boldsymbol{H}_{\boldsymbol{q},i}^{\text{ext}} \tag{20}$$

$$\boldsymbol{H}_i^{\text{int}} = \frac{\partial \boldsymbol{h}_i}{\partial \boldsymbol{r}_i^C} = \begin{pmatrix} \frac{f_x}{z_i^C} & 0 & -\frac{f_x}{(z_i^C)^2} x_i^C \\ 0 & \frac{f_y}{z_i^C} & -\frac{f_y}{(z_i^C)^2} y^C \end{pmatrix} \tag{21}$$

$$\boldsymbol{H}_{\boldsymbol{q},i}^{\text{ext}} = \frac{\partial \boldsymbol{r}_i^C}{\partial \boldsymbol{q}} = \frac{\partial \left(\boldsymbol{q} \otimes \boldsymbol{r}_i^B \otimes \boldsymbol{q} \right)}{\partial \boldsymbol{q}}; \quad \boldsymbol{H}_{\boldsymbol{t}^C,i}^{\text{ext}} = \frac{\partial \boldsymbol{r}_i^C}{\partial \boldsymbol{t}^C} = \boldsymbol{I}_3. \tag{22}$$

The partial derivative in Eqn.22 is computed according to Ref. [24].

After the measurements update step, brute-force normalization is performed to ensure unity of the estimated quaternion [24].

VII. Simulations

In this section, the simulation environment and the results are presented. Firstly, the impact of including a heatmaps-derived covariance in the pose estimation step is addressed by comparing the CEPPnP method with a standard solver which does not account for feature uncertainty. Secondly, the performance of the EKF is evaluated by comparing the convergence profiles with a heatmaps-derived covariance matrix against covariance matrices with arbitrary selected covariances. Initialization is provided by the CEPPnP for all the scenarios.

Two separate error metrics are adopted in the evaluation, in accordance with Ref. [11]. Firstly, the translational error between the estimated relative position $\hat{\boldsymbol{t}}^C$ and the ground truth $\boldsymbol{t}$ is computed as

$$E_T = |\boldsymbol{t}^C - \hat{\boldsymbol{t}}^C|. \tag{23}$$

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

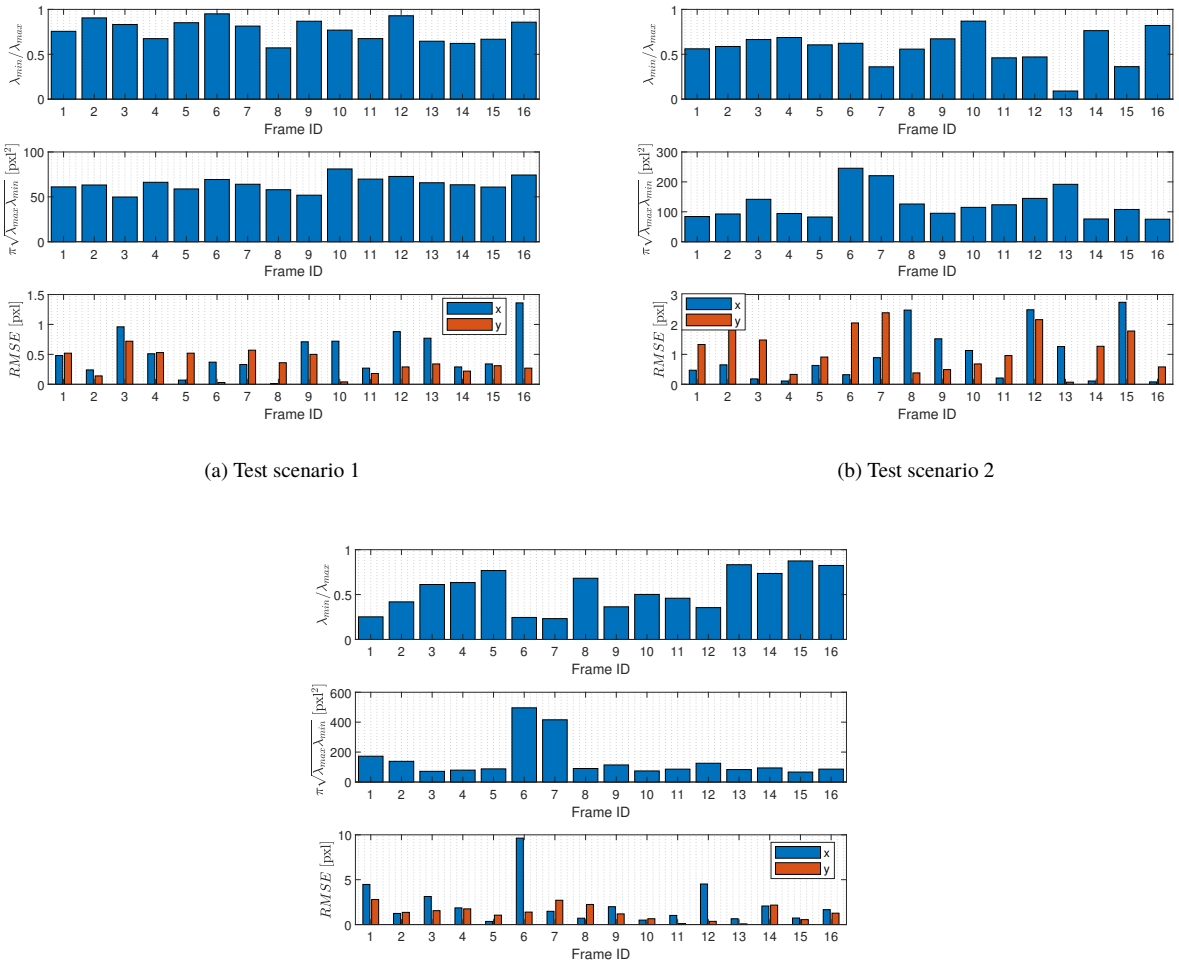
$$\boldsymbol{\beta} = \begin{pmatrix} \beta_s & \boldsymbol{\beta_v} \end{pmatrix} = \boldsymbol{q} \otimes \hat{\boldsymbol{q}} \tag{24}$$

$$E_R = 2 \arccos\left(|\beta_s|\right). \tag{25}$$

A. Pose Estimation

Three representative scenarios are selected from the CNN test dataset for the evaluation. These scenarios were chosen in order to analyze different heatmaps' distributions around the detected features. A comparison is made between the proposed CEPPnP and the EPPnP. Figure 8 shows the characteristics of the covariance matrices derived from the predicted heatmaps. Here, the ratio between the minimum and maximum eigenvalues of the associated covariances is represented against the ellipse's area and the RMSE between the Ground Truth (GT) and the x, y coordinates of the extracted features,

$$E_{\text{RMSE},i} = \sqrt{(x_{GT,i} - x_i)^2 + (y_{GT,i} - y_i)^2}. \tag{26}$$



(a) Test scenario 1

(b) Test scenario 2

(c) Test scenario 3

Fig. 8 Characteristics of the ellipses derived from the covariance matrices for the three selected scenarios.

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

Table 2 Pose Estimation performance results.

Metric	Scenario	CEPPnP	EPPnP
E_T [m]	1	[0.18 0.22 0.24]	[0.17 0.22 0.24]
	2	[0.35 0.41 0.59]	[0.14 0.4 22.8]
	3	[0.49 0.12 1.41]	[0.56 0.16 5.01]
E_R [deg/s]	1	0.36	0.35
	2	0.75	6.08
	3	1.99	2.72

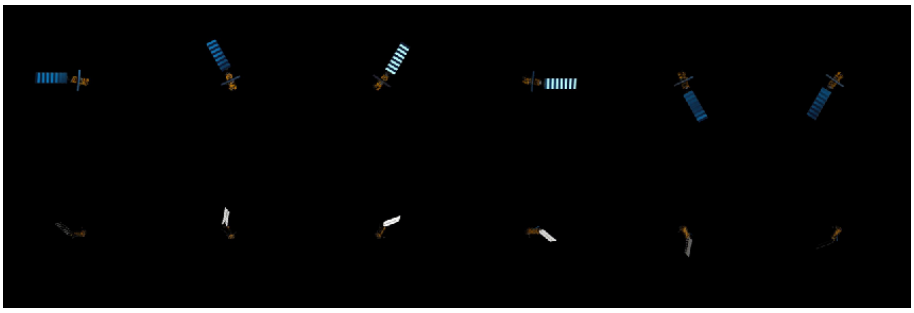


Fig. 9 Montage of the VBAR approach scenario 1 (Top) and 2 (Bottom). Images are shown every 12 s for clarity.

Table 2 lists the pose estimation results for the three scenarios. As anticipated in Figure 8, the statistical information derived from the heatmaps in the first scenario is uniform for all the features, due to the very accurate CNN detection. As a result, the inclusion of features covariance in the CEPPnP solver does not help refining the estimated pose. Both solvers are characterized by the same pose accuracy.
Not surprisingly, the situation changes as soon as the heatmaps are not uniform across the feature IDs. Due to its capability of accomodating feature uncertainties in the estimation, the CEPPnP method outperforms the EPPnP for the remaining scenarios. In other words, the CEPPnP solver proves to be more robust against inaccurate CNN detections by accounting for a reliable representation of the features covariance.

B. Navigation Filter

To assess the performance of the proposed EKF, two rendezvous scenarios with Envisat are rendered in Cinema 4D©. These are perturbation-free VBAR trajectories characterized by a relative velocity $\|\boldsymbol{v}\| = 0$ m/s. In both scenarios, the Envisat performs a roll rotation of $\|\boldsymbol{\omega}\| = 5$ deg/s, with the servicer camera frame aligned with the LVLH frame. Table 3 lists the initial conditions of the trajectories, whereas Figure 9 shows some of the associated rendered 2D images. It is assumed that the images are made available to the filter every 2 seconds for the measurement update step, with the propagation step running at 1 Hz. In both scenarios, the EKF is initialized with the CEPPnP pose solution at time t_0. The other elements of the initial state vector are randomly chosen assuming a standard deviation of 1 mm/s and 1 deg/s for all the axes of terms $(\hat{\boldsymbol{v}}_0 - \boldsymbol{v})$ and $(\hat{\boldsymbol{\omega}}_0 - \boldsymbol{\omega})$, respectively. Table 4 reports the initial conditions of the filter.

Table 3 VBAR approach scenarios. The attitude is represented in terms of ZYX Euler angles for clarity.

Scenario	$\boldsymbol{\theta}_0$ [deg]	$\boldsymbol{\omega}$ [deg/s]	$\boldsymbol{t}^C$ [m]	$\boldsymbol{v}$ [mm/s]
1	$[-180\ 30\ -80]^T$	$[-2.5\ -4.3\ 0.75]^T$	$[0\ 150\ 0]^T$	$[0\ 0\ 0]^T$
2	$[-180\ -50\ 160]^T$	$[3.8\ 1.1\ -3.02]^T$	$[0\ 180\ 0]^T$	$[0\ 0\ 0]^T$

In order to assess the impact of the measurements covariance on the performance of the EKF, two diagonal matrices are considered aside from the proposed heatmaps-derived covariance in Eqn.15: a matrix $\boldsymbol{R} = 10 \cdot \boldsymbol{I}_{2n}$, in which 10 is the

Table 4 Initial state vector in the EKF.

Scenario	$\hat{\boldsymbol{\theta}}_0$ [deg]	$\hat{\boldsymbol{\omega}}_0$ [deg/s]	$\hat{\boldsymbol{t}}_0^C$ [m]	$\hat{\boldsymbol{v}}_0$ [mm/s]
1	[-180.17 28.74 -80.63]^T	[-2.07 -4.1 0.1]^T	[0.11 0.07 149.69]^T	[2.8 -1.3 3]^T
2	[-179 -51.06 156.25]^T	[-2.07 -4.1 0.1]^T	[0.10 0.26 181.9]^T	[2.8 -1.3 3]^T

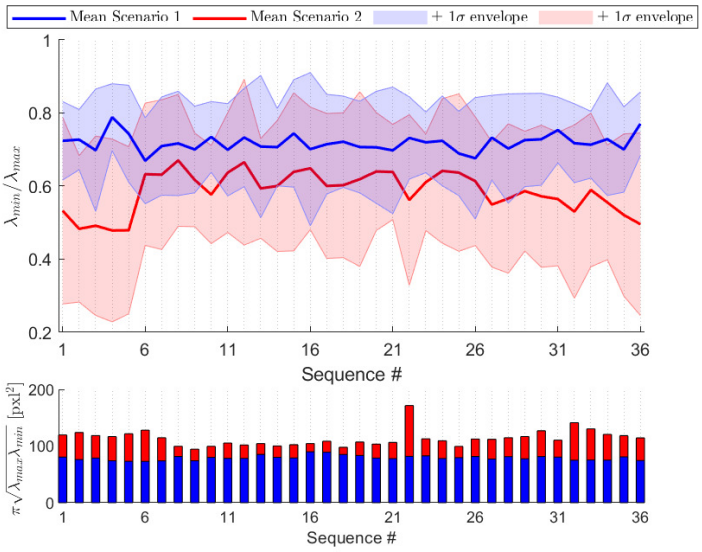


Fig. 10 Comparison of the selected navigation scenarios throughout the whole rotational sequence of Envisat, based on the mean eccentricity (Top) and mean area (Bottom) derived from the features covariances.

selected variance, and a matrix $\boldsymbol{R}_{\text{avg}}$, which uses an averaged variance of the diagonal elements of the heatmaps-derived covariance.

Figure 10 illustrates two selected mean characteristics of the heatmaps, calculated over the n features in each frame, for the two rotational sequences of the Envisat. First of all, the mean eccentricity of the feature covariances, represented by the ratio $\lambda_{min}/\lambda_{max}$, indicates that the first scenario is characterized by more circular heatmaps than the second one. Moreover, the intra-frame dispersion of the ellipses' shape, represented by the $\pm 1\sigma$ envelope, is much larger for the second scenario than it is for the first one. This suggests that the covariances $\boldsymbol{C}_i$ in the second scenario can be very different from each other within a given frame. Lastly, it can be seen that the ellipses' areas in the second scenario are larger than the ones in the first scenario throughout the entire rotational sequence. Overall, these considerations emphasize a worse detection accuracy for certain features in the second scenario.

Figures 11-12 show the convergence profiles for the translational and rotational state for the first scenario, respectively. Overall, the selection of the measurements covariance seems not to impact on the filter performance. As already anticipated, this is a consequence of the nearly circular, small heatmaps which characterize almost every feature throughout this sequence. As such, a diagonal covariance of 10 pixels, an heatmap-derived covariance, or an averaged diagonal covariance return an almost identical convergence profile. The situation differs considerably when analyzing the convergence profiles for the second scenario (Figures 13-14). Despite similar convergences occur for the rotational state as well as for the relative velocity, it can be seen that the selection of the measurement covariance plays a key role in the along-track error. Specifically, the error trend of the heatmaps-derived covariance oscillates around zero as opposed to the other two biased estimates. This seems to indicate that the statistical information derived from the CNN heatmaps can represent the measurements uncertainty and aid the filter convergence also in case of inaccurate feature detections.

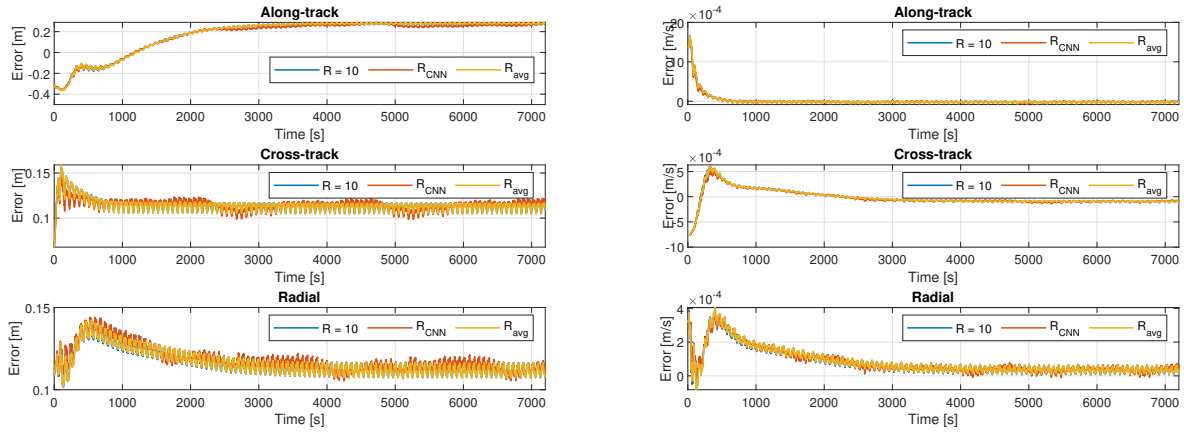


Fig. 11 Estimation error for the relative position and velocity - Scenario 1.

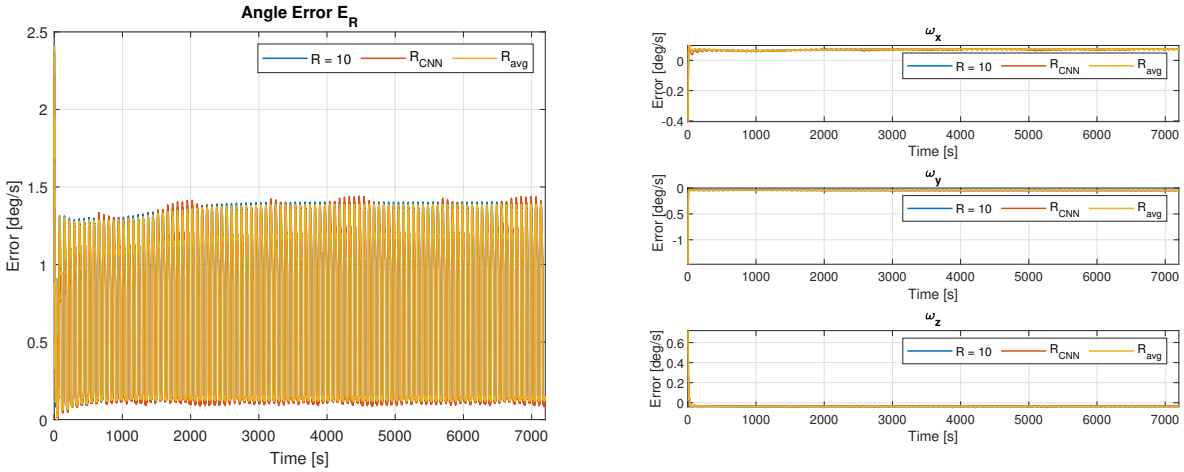


Fig. 12 Estimation error for the relative attitude and angular velocity - Scenario 1.

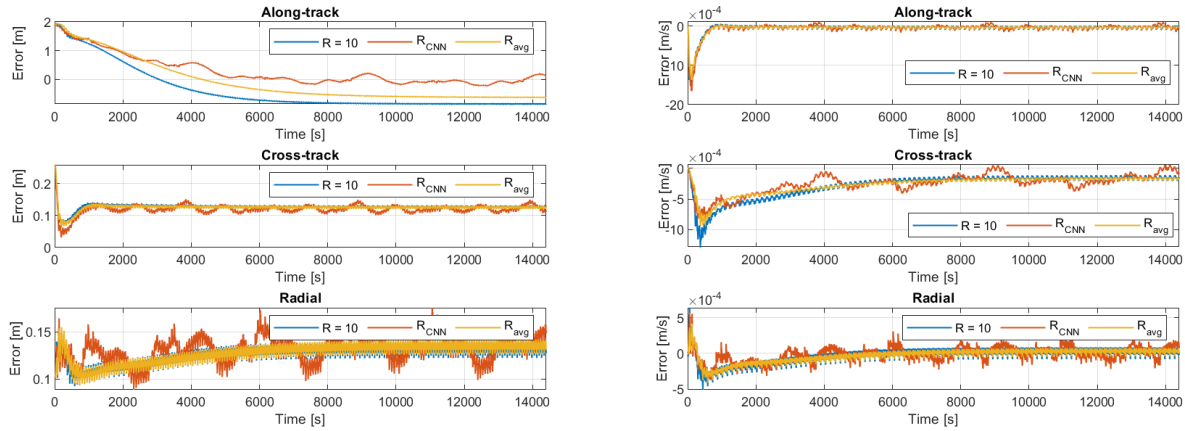


Fig. 13 Estimation error for the relative position and velocity - Scenario 2.

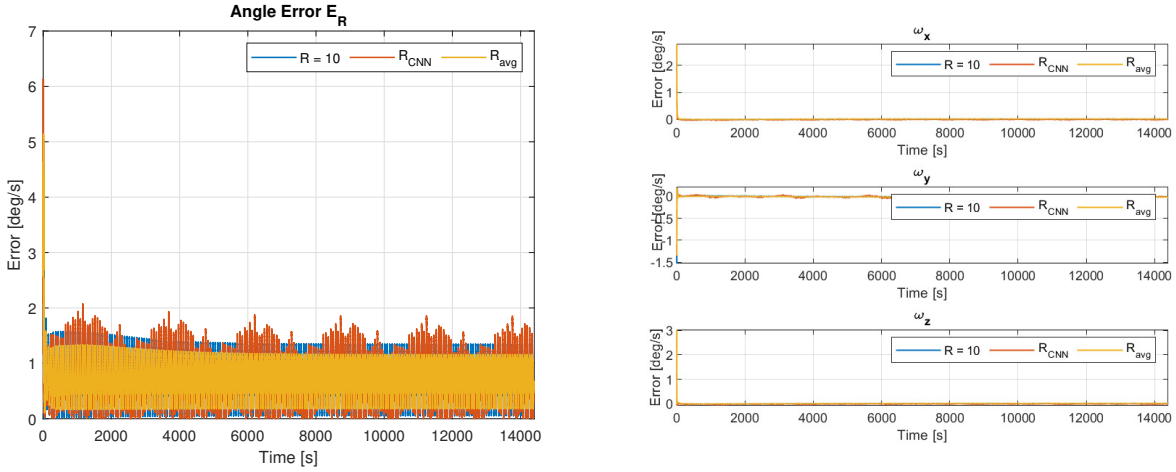


Fig. 14 Estimation error for the relative attitude and angular velocity - Scenario 2.

VIII. Conclusions and Recommendations

This paper introduces a novel framework to estimate the relative pose of an uncooperative target spacecraft with a single monocular camera onboard a servicer spacecraft. A method is proposed in which a CNN-based IP algorithm is combined with a CEPPnP solver and an EKF to return a robust estimate of the relative pose as well as of the relative translational and rotational velocities.

The main novelty of the proposed CNN-based method stands in introducing a heatmaps-derived covariance representation of the detected features. This is done in order to incorporate feature uncertainties in both the pose estimation step and in the navigation filter. Preliminary results indicate that the correlation between the accuracy of the CNN detection and the shape of the detected heatmaps can be exploited to compute a covariance matrix for each detected feature. This covariance can be incorporated in the CEPPnP solver to improve the pose estimation accuracy under inaccurate CNN detections. Furthermore, the same statistical information can be used to build a representative measurements covariance matrix and improve the measurements update step of the navigation filter. Together, these improvements proved to return a more robust and accurate estimate for the considered VBAR approach scenarios. Overall, the results suggest a promising scheme to cope with the challenging demand for robust navigation in close-proximity scenarios.

However, further work is required in several directions. First of all, more recent CNN architectures shall be investigated to assess the achievable robustness and accuracy in the feature detection step. Secondly, the comparison between the CEPPnP and the covariance-free EPPnP solvers shall be extended to the entire test dataset, in order to validate the improved accuracy so far observed for only three representative scenarios. Moreover, the NRM shall be included in the pose estimation step to refine the accuracy of the CEPPnP solver and improve filter initialization. Besides, more close-proximity scenarios shall be recreated to investigate more complex relative trajectories as well as to assess the impact of perturbations on the accuracy and robustness of the navigation filter. In this context, other navigation filters such as the Unscented Kalman Filter and the Multiplicative EKF shall be investigated. Finally, alternative datasets such as SPEED shall be considered as a benchmark for the proposed method.

Acknowledgments

This study is funded and supported by the European Space Agency and Airbus Defence and Space under Network/Partnering Initiative (NPI) program with grant number NPI 577 – 2017.

References

[1] Tatsch, A., Fitz-Coy, N., and Gladun, S., "On-orbit Servicing: A Brief Survey," *Proceedings of the 2006 Performance Metrics for Intelligent Systems Workshop*, 2006, pp. 21–23.

[2] Wieser, M., Richard, H., Hausmann, G., Meyer, J.-C., Jaekel, S., Lavagna, M., and Biesbroek, R., "e.Deorbit Mission: OHB Debris Removal Concepts," *ASTRA 2015-13th Symposium on Advanced Space Technologies in Robotics and Automation*, Noordwijk, The Netherlands, 2015.

[3] Sharma, S., Ventura, J., and D'Amico, S., "Robust Model-Based Monocular Pose Initialization for Noncooperative Spacecraft Rendezvous," *Journal of Spacecraft and Rockets*, Vol. 55, No. 6, 2018, pp. 1–16.

[4] Pasqualetto Cassinis, L., Fonod, R., and Gill, E., "Review of the Robustness and Applicability of Monocular Pose Estimation Systems for Relative Navigation with an Uncooperative Spacecraft," *Progress in Aerospace Sciences*, Vol. 110, 2019.

[5] Lepetit, Moreno-Noguer, F., and Fua, P., "EPnP: an accurate O(n) solution to the PnP problem," *International Journal of Computer Vision*, Vol. 81, 2009, pp. 155–166.

[6] Ferraz, L., Binefa, X., and Moreno-Noguer, F., "Very Fast Solution to the PnP Problem with Algebraic Outlier Rejection," *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014.

[7] Ostrowsky, A., *Solution of Equations and Systems of Equations*, 2nd ed., Academic Press, New York, 1966.

[8] Ferraz, L., Binefa, X., and Moreno-Noguer, F., "Leveraging Feature Uncertainty in the PnP Problem," *Proceedings of the British Machine Vision Conference*, Nottingham, UK, 2014.

[9] Cui, J., Min, C., Bai, X., and Cui, J., "An Improved Pose Estimation Method Based on Projection Vector with Noise Error Uncertainty," *IEEE Photonics Journal*, Vol. 11, No. 2, 2019.

[10] Sharma, S., Beierle, C., and D'Amico, S., "Pose Estimation for Non-Cooperative Spacecraft Rendezvous using Convolutional Neural Networks," *IEEE Aerospace Conference*, Big Sky, MT, USA, 2018.

[11] Sharma, S., and D'Amico, S., "Pose Estimation for Non-Cooperative Spacecraft Rendezvous using Neural Networks," *29th AAS/AIAA Space Flight Mechanics Meeting*, Ka'anapali, HI, USA, 2019.

[12] Shi, J., Ulrich, S., and Ruel, S., "CubeSat Simulation and Detection using Monocular Camera Images and Convolutional Neural Networks," *2018 AIAA Guidance, Navigation, and Control Conference*, Kissimmee, FL, USA, 2018.

[13] Rondao, D., and Aouf, N., "Multi-View Monocular Pose Estimation for Spacecraft Relative Navigation," *2018 AIAA Guidance, Navigation, and Control Conference*, Kissimmee, FL, USA, 2018.

[14] Capuano, V., Alimo, S., Ho, A., and Chung, S., "Robust Features Extraction for On-board Monocular-based Spacecraft Pose Acquisition," *AIAA Scitech 2019 Forum*, San Diego, CA, USA, 2019.

[15] Pavlakos, G., Zhou, X., Chan, A., Derpanis, K., and Daniilidis, K., "6-DoF Object Pose from Semantic Keypoints," *IEEE International Conference on Robotics and Automation*, 2017.

[16] Newell, A., Yang, K., and Deng, J., "Stacked Hourglass Networks for Human Pose Estimation," *European Conference on Computer Vision*, Springer, 2016, pp. 483–499.

[17] Chen, B., Cao, J., Parra, A., and Chin, T., "Satellite Pose Estimation with Deep Landmark Regression and Nonlinear Pose Refinement," *International Conference on Computer Vision*, Seoul, South Korea, 2019.

[18] Park, T., Sharma, S., and D'Amico, S., "Towards Robust Learning-Based Pose Estimation of Noncooperative Spacecraft," *AAS/AIAA Astrodynamics Specialist Conference*, Portland, ME, USA, 2019.

[19] D'Amico, S., Benn, M., and Jorgensen, J., "Pose Estimation of an Uncooperative Spacecraft from Actual Space Imagery," *International Journal of Space Science and Engineering*, Vol. 2, No. 2, 2014, pp. 171–189.

[20] Curtis, H., *Orbital Mechanics for Engineering Students*, Elsevier, 2005.

[21] Ronneberger, O., Fischer, P., and Brox, T., "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 234–241.

[22] Pasqualetto Cassinis, L., Fonod, R., Gill, E., Ahrns, I., and Gil Fernandez, J., "Comparative Assessment of Image Processing Algorithms for the Pose Estimation of an Uncooperative Spacecraft," *International Workshop on Satellite Constellations & Formation Flying*, Glasgow, UK, 2019.

[23] Sharma, S., and D'Amico, S., "Reduced-Dynamics Pose Estimation for Non-Cooperative Spacecraft Rendezvous Using Monocular Vision," *Advances in the Astronautical Sciences Guidance, Navigation and Control*, Vol. 159, 2017.

[24] Civera, J., Davison, A., and Martínez Montiel, J., *Structure from Motion using the Extended Kalman Filter*, Springer, 2012.