A robust approach merging deep learning and unscented Kalman for vision based space rendez-vous

Thomas Chambon, Fabien Gallet

75th International Astronautical Congress (IAC), Milan, Italy, 14-18 October 2024.

IAC–24–IAC-24,C1,IPB,27,x85292

A robust approach merging deep learning and unscented Kalman for vision based space rendez-vous

Thomas Chambon[a*], **Fabien Gallet**[b]

[a] *IRT Saint-Exupéry, 240 Rue Evariste Galois, 06410 Biot (France), thomas.chambon@irt-saintexupery.com*
[b] *IRT Saint-Exupéry, 240 Rue Evariste Galois, 06410 Biot (France), fabien.gallet@irt-saintexupery.com*
[*] *Corresponding author*

Abstract

Autonomous space rendezvous (RDV) is gaining increasing significance in tandem with the advancement of on-orbit services and operations, such as refueling, debris removal, in-orbit assembly, and transportation. Several of these applications require the ability to navigate around non-prepared clients, where specific navigation aids are unavailable.

In this context, vision-based navigation approaches present cost-effective solutions. Furthermore, the recent strides in deep learning methodologies have enhanced the robustness and performance of image processing. The considerable advances in on-board computation capabilities, coupled with the concurrent development of deep learning deployment chains, allows the on-board integration of such approaches. These progresses open up the opportunity of incorporating deep learning-based vision approaches in navigation solutions for satellite rendezvous missions.

This paper presents a navigation solution developed with the objective of conceiving an on-board sub-system able to plan and realize autonomously a rendezvous mission. The conceived end-to-end navigation approach combines the use of an Artificial Neural Network (able to estimate the positions of keypoints in the processed images) and of an Unscented Kalman filter (able to bring the knowledge of both geometry and dynamics to provide relative pose estimation). Particular attention has been paid to develop a robust solution suitable for an operational use and on-board processor deployment.

Exhaustive efforts have been made to formulate a robust initialization algorithm to ensure the convergence of navigation. This processing step is indeed critical in an operational environment and constitutes the most important contribution of this deep learning solution. To improve precision and detect outliers, a specialized interface between the Artificial Neural Network and the Unscented Kalman filter is proposed. Finally, the performance of the solution has been thoroughly evaluated using a realistic dataset of approach sequences.

Keywords: Space rendezvous, Image based navigation, Pose estimation, GNC

Nomenclature

$\Delta^z_{i,k}$	Filter measurement innovation at timestep k for keypoint i
Ω	Client rotation matrix with respect to Local Orbital Frame
ω	Client rotation velocity vector in client frame
ω_{orb}	Orbital pulsation
Σ	Covariance matrix for keypoints (2×2)
H	Heatmaps estimated by the neural network
M_i	Keypoint coordinates in client 3D model
p	Modified Rodrigues parameters representation
P_{CAM}	Camera projection matrix
q	Quaternion representation
R	Rotation matrix from client frame to camera frame given in camera frame
S_k	Filter measurement covariance at time step k
t	Translation client frame to camera frame given in camera frame

Acronyms/Abbreviations

ADR Active Debris Removal. 2
AOCS Attitude and Orbit Control System. 6, 8
CNN Convolutional Neural Networks. 2, 3, 10, 11
CWS Clohessy-Wiltshire. 6
EKF Extended Kalman Filter. 2
IOS In-Orbit Servicing. 2
IP Image Processing. 9
LOF Local Orbital Frame. 5, 6, 8
NAC Narrow Angle Camera. 3
PnP Perspective-n-Points. 3
PSD Power Spectral Density. 9

UKF Unscented Kalman Filter. 2, 3, 8, 10, 11

1. Introduction

With the rise of In-Orbit Servicing (IOS) missions focused on Active Debris Removal (ADR), satellite lifespan extension, refueling, and repairing, mastering space rendezvous has become essential. Space rendezvous demands both high safety standards to meet stringent debris limitation requirements as well as autonomy due to limited ground-to-satellite communication. Consequently, developing safe, standardized, and cost-effective solutions is crucial. Visual based pose estimation and tracking for relative navigation seems promising but remains challenging for a unprepared client. By "unprepared", it is meant that the client satellite has no specific mechanical, communication or navigation interface for rendezvous. However, its 3D model is known. The main difficulties related to visual based navigation are:

- To provide a robust solution compliant with collision risk mitigation requirements.
- To demonstrate capability to manage domain gap between laboratory or simulated environment and real space environment that affects image content.
- To offer a solution that does not require direct communication between servicer and client satellites.
- To develop algorithms compatible with real time execution on on-board computer

Historically, most efficient conventional algorithms for pose estimation relied on Model-to-image matching architectures [1] where handcrafted features were extracted from image and matched with model representations [2, 3]. Many of these solutions are iterative tracking approaches that require a good prior. Although tracking approaches yield good results, they are difficult to initialize and are prone to drift or divergence when the initial estimate is incorrect. Additionally, handcrafted features lack robustness against variations in background or lighting conditions.

Over the last years, Convolutional Neural Networks (CNN) architectures have demonstrated their advantages for image interpretation. They offer state of the art performances for pose estimation problems with good robustness properties [4, 5]. However, their behaviour can be hardly predicted and their integration into a real time and critical close loop system is still a challenge to demonstrate proper robustness and fault identification.

Initially, research focused on directly estimating the relative pose of the client as a 6D vector at the network's output [6, 7]. However, the behavior of these architectures is difficult to interpret, and image processing does not benefit from the information that would be provided by a navigation filter to improve pose estimation. State-of-the-art pose estimation generally relies on dense maps for pixel matching [8–10], but this typically requires a significant computational load.

Allowing tighter coupling between image processing and navigation filter, solutions with a reduced number of 2D keypoints at the output of the network have been proposed. These works have been inspired by architectures initially developed for human pose estimation [11–13].

Image processing outputs need to be filtered out taking into account temporal dynamic to reach the proper level of precision and to improve robustness. As suggested in [14, 15], tightly coupled approaches that directly interface with 2D localized features in the image plane are more performant and robust. These methods are also more versatile, allowing for potential fusion with other measurements at the filter level. The main challenge with this approach is the non linearity of the image measurement model, which requires the use of Extended Kalman Filter (EKF) or Unscented Kalman Filter (UKF).

Having a robust estimator of measurement uncertainty is one of the most important features to construct a robust solution. In [15, 16], the authors have developed a post-processing method to estimate keypoint covariance from heatmaps at the output of the CNN. However, such handcrafted method can suffer from limitations when heatmap shapes derive from learnt gaussian model.

This work is focused on the integration of a CNN image processing solution for keypoint estimation with a Kalman navigation filter applied to the last force manoeuvres before docking (last 200 meters). The main contributions are the following:

1. The development of a specific image processing solution based on CNN for keypoint regression
2. The implementation of a tightly coupled Unscented Kalman Filter (UKF) integrating image measurements for collaborative rendezvous
3. The development of a specific CNN head at the interface with UKF to estimate the measurement

covariance

4. The demonstration of the end-to-end navigation solution performances over a dedicated dataset of sequences

2. Navigation overall architecture

The proposed architecture combines a deep neural network and a UKF filter. The CNN takes the gray level image of fixed dimensions 512×512 as input and outputs keypoints coordinates associated to their covariance estimates.

The navigation filter propagates the kinematic and orbital dynamics and updates its state based on the the output from the CNN. The filter is designed for cooperative rendezvous scenarios where the client attitude is supposed to be controlled close to its nominal position.

Since keypoint measurement model and kinematic propagation model are non linear with respect to state parameters, an Unscented Kalman Filter implementation is used.

A thightly coupled approach is chosen to interface the output of the image processing with the navigation filter. Such an approach, as compared with a Perspective-n-Points (PnP) (see [17]) algorithm directly at the output of the image processing has the following advantages:

- Possibility to fuse complementary information from other sensors such as range measurements at filter level
- An increased robustness at Kalman filter level by identifying bad keypoints. This filtering can benefit from the dynamic and kinematic model knowledge or other sensors redundant information.

To get optimal performances, the input image is cropped around the estimated position of the client (see diagram 1). This crop is initialized at first image through an iterative process, and then propagated by the navigation filter to resize the next image samples.

3. Image processing architecture

3.1 Image dataset

The navigation solution has been trained, tuned and tested using two high fidelity datasets. A first dataset of singleton images was dedicated to the training of the image processing neural network. It consists of a 120000 samples dataset generated with Thales Alenia Space ©**SPICaM** software. It is specifically designed for spacecraft pose estimation ensuring fine statistical balancing and realistic dynamic ranges using a complete camera model. The camera model contains a physically based ray-tracer coupled with a complete sensor model taking into account key camera limitations. An additional test split is provided for testing the robustness of the proposed solution. This dataset has been released and more details are discussed in the related publication [18].

A second dataset of 160 random approach sequences has been generated to test the full navigation solution. The simulations cover trajectories from 200 to 30m, corresponding to the force motion domain. The limitation at 30m is due to the simulated Narrow Angle Camera (NAC), but a similar behaviour is expected at lower distance. Image generation uses the **SPICaM** based rendering tool chain coupled with a high fidelity numerical orbit propagator. The orbit propagation is handled using a fine orbit propagator based on Orekit (Orbit Extrapolation Kit) and takes into account non linear effects and higher order perturbations such as J2. Spacecraft actuators limitations and uncertainties are also modelled.

The figure 2 shows the servicer directions viewed from the client frame for all the generated sequences.

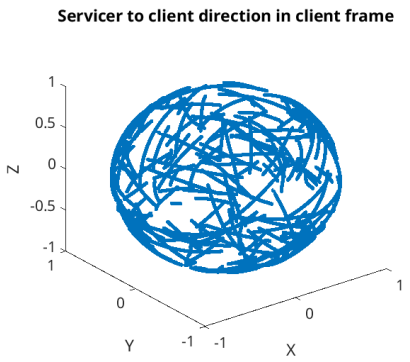


Fig. 2. Servicer directions expressed in client frame

3.2 Deep learning architecture

Keypoints placed on the spacecraft are detected on the images using a Convolutional Neural Networks (CNN). The figure 3 summarizes its architecture. This model has been designed to be compatible with real time applications. The backbone is composed of a MobileNet [19] network coupled with a bi-directional features pyramidal network [20]. The prediction head is built with few successive convolutional layers, and delivers for each keypoint an intermediate output that can be considered as a heatmap **H**.

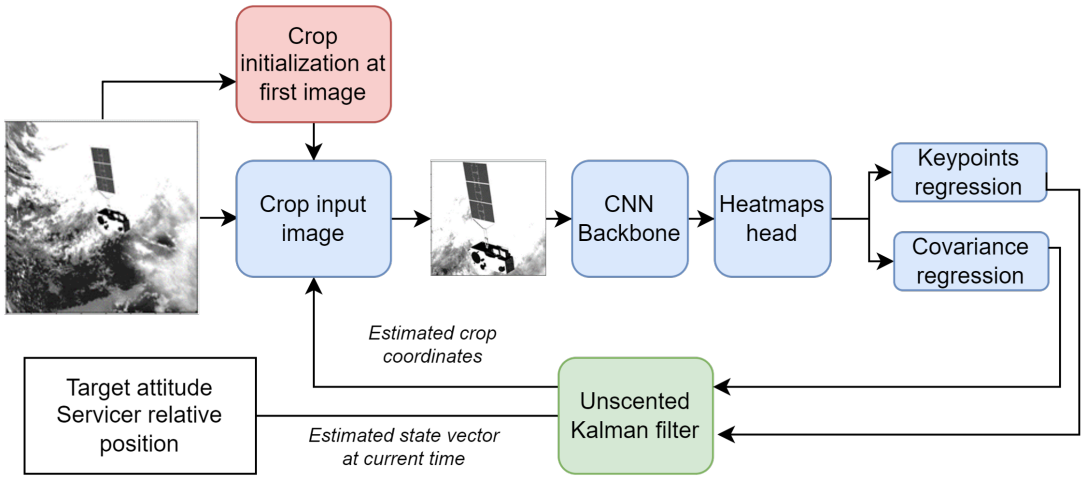


Fig. 1. Diagram of overall navigation architecture

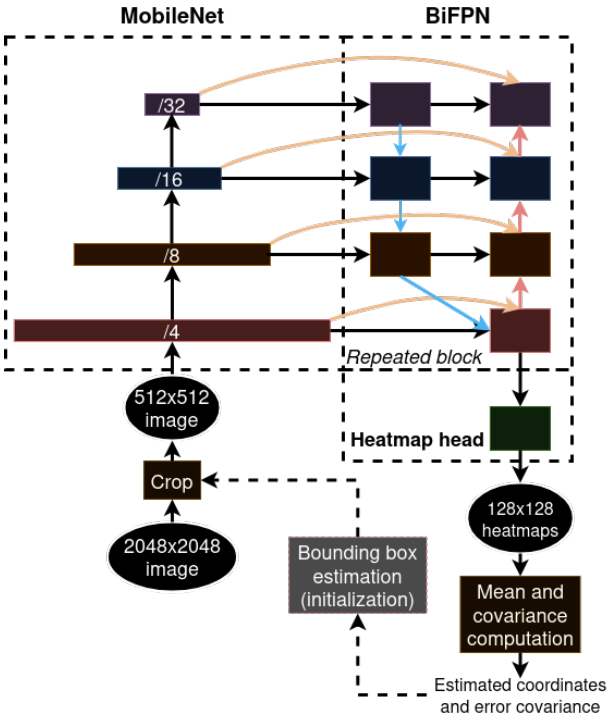


Fig. 3. Image processing architecture overview

3.3 Keypoint estimation

The resulting heatmap $\mathbf{H}$ is viewed as the probability for the keypoint to be present at each pixel. The keypoint coordinate is regressed as the mean of the resulting distribution.

$$\begin{bmatrix}\hat{x}\\ \hat{y}\end{bmatrix} = \frac{1}{\sum_{\{i,j\}=1}^{N}\mathbf{H}(i,j)}\begin{bmatrix}\sum_{\{i,j\}=1}^{N}\mathbf{H}(i,j)i\\ \sum_{\{i,j\}=1}^{N}\mathbf{H}(i,j)j\end{bmatrix} \quad [1]$$

3.4 Covariance estimation

Estimating uncertainty at the output of the network is key for providing a robust solution and identifying erroneous measurements at the entrance of the navigation filter. Uncertainty can be either epistemic, refering to model uncertainty or aleatoric when associated to database uncertainty distributions. This work is focused at estimating epistemic uncertainty at the output of the network.

Similarly to keypoint head, covariance can be regressed from heatmap by the following differentiable function (equation 2) where: $\mathbf{H}(i,j)$ represents the heatmap value at position (i,j).

$$\hat{\mathbf{\Sigma}} = \frac{1}{\sum_{\{i,j\}=1}^{N}\mathbf{H}(i,j)}\begin{bmatrix}\Sigma_{\{i,j\}=1}^{N}\mathbf{H}(i,j)i^2 & \Sigma_{\{i,j\}=1}^{N}\mathbf{H}(i,j)ij\\ \Sigma_{\{i,j\}=1}^{N}\mathbf{H}(i,j)ij & \Sigma_{\{i,j\}=1}^{N}\mathbf{H}(i,j)j^2\end{bmatrix} \quad [2]$$

Inspired by [21], the covariance is estimated at the output of the neural network in addition to keypoint regressed coordinates through negative log likelyhood formulation under gaussian hypothesis. Recall that under gaussian assumption, the probability of having x assuming mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Sigma}$ is written as per equation 3.

$$f(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}} \quad [3]$$

Then, the heatmap loss can be expressed as the negative log-likelyhood as per equation 4.

$$\begin{aligned}\mathcal{L}_{hm}(\mathbf{x}, \hat{\mathbf{x}}, \hat{\boldsymbol{\Sigma}}) &= -\log f\left(\mathbf{x}, \hat{\mathbf{x}}, \hat{\boldsymbol{\Sigma}}\right) \\ &= \frac{1}{2}(\mathbf{x}-\hat{\mathbf{x}})^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}-\hat{\mathbf{x}}) + \log\left(|\hat{\boldsymbol{\Sigma}}|\right) \\ &+ const.\end{aligned} \quad [4]$$

Through this loss, the shape of the heatmap is constrained to model both the uncertainty and the average value (see figure 4). This approach has demonstrated good results even under severe domain gap.

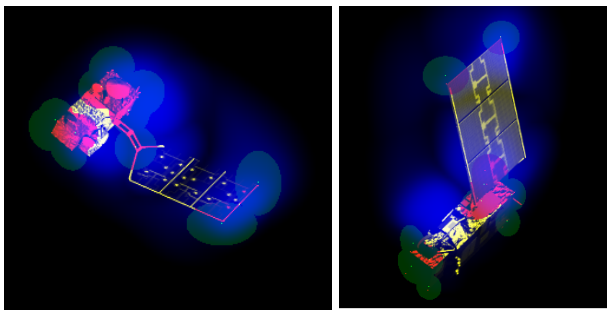


Fig. 4. Example of heatmaps and covariance estimation. Heatmaps are plotted in blue while covariance are superimposed in green.

The accuracy and the covariance estimation are studied jointly to evaluate the performances of the image processing neural network. Figure 5 shows the consistency between the error and its estimated covariance. Outliers represent less than 1% of the estimated data.

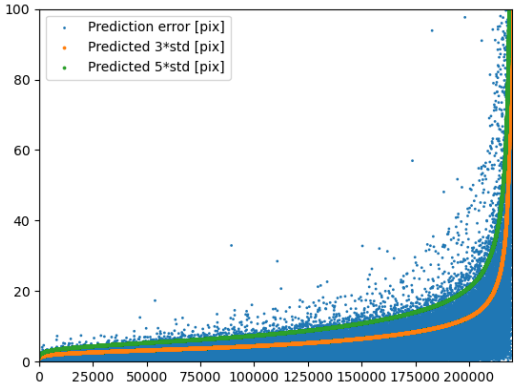


Fig. 5. Scatter plot of keypoint error norms, sorted by ascending estimated standard deviation

3.5 Crop initialization

The model is trained to take a cropped image as input, with the spacecraft covering the majority of the image. Nevertheless, without an initial guess of the bounding box on the full image, it is possible to use the region enclosing all of the keypoints uncertainties ellipses as a gross prediction of the spacecraft bounding box.

The initialization process is performed using a simple loop, where each iteration adjusts the cropping area. Keypoint estimated covariances reduce thanks to cropping. Typically, the loop requires three iterations to identify an approximate cropping box that aligns with the training domain. Figure 6 shows an example of initialization sequence.

4. Kalman filter

4.1 Orbit propagation model

All further equations are derived in the Local Orbital Frame (LOF) defined as in figure 7.

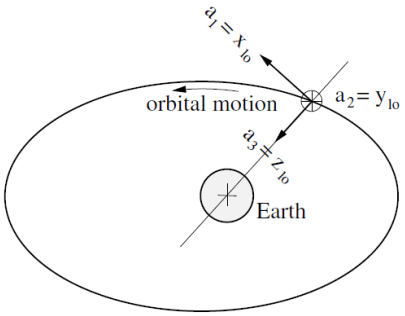


Fig. 7. Definition of Local Orbital Frame (extract from [22])

Fig. 6. Initialization sequence, from the left to the right, the yellow box represents the estimated cropping box used on the next frame

For the presented cooperative usecase, during the last step of the approach in force motion, client and servicer are very close to each other (below 200m) and assumed on an circular orbit. Therefore one can consider that the traditionnal first order approach corresponding to Clohessy-Wiltshire (CWS) will be sufficient to capture the orbital dynamic neglecting minor perturbations such as J_2 and drag. Indeed, the CWS approximation holds under the following hypotheses:

- The distance between the servicer and the client is small compared to the orbital radius
- The orbit is nearly circular

$$\begin{aligned}\ddot{x} &= 2\omega_{orb}\dot{z} + \frac{1}{m_c}F_x \\ \ddot{y} &= -\omega_{orb}^2 y + \frac{1}{m_c}F_y \\ \ddot{z} &= -2\omega_{orb}\dot{x} + 3\omega_{orb}^2 z + \frac{1}{m_c}F_z \end{aligned} \quad [5]$$

$\omega_{orb} = \frac{2\pi}{T}$ is the orbit angular frequency and F_x, F_y, F_z are the forced applied to the servicer.

Works have also been proposed to tackle the problem of elliptical orbits by linearizing the equations around the nominal elliptical orbit. These equations, as derived by Tschauner and Hempel [23] can be easily adapted to the proposed filter architecture.

Close form solutions can be obtained for this set of linear equations. The derived state transition matrices are used to propagate orbit parameters taking input on-board knowledge of thrust forces applied on servicer (see [22]).

4.2 Attitude propagation model

The client being cooperative, only small attitude deviations with respect to its nominal orientation in the LOF are estimated to increase filter robustness and detect any client AOCS failure. For what regards servicer attitude knowledge, star-trackers inertial information will be one order of magnitude more precise than image processing information. It is therefore used as an input to the filter.

Since only small deviations are assumed with respect to nominal client attitude, modified Rodrigues Parameters (MRP) are used as a 3 dimension representation. A good overview of these parameterizations, their advantages and drawbacks is given in [24]. MRP is singular for rotations of +/-360°, far enough to be used in the update of the filter with small increments.

$$p = \frac{1}{1+q_0} q_{1:3} \quad [6]$$

Note that in [16] and [25] p is defined as $4p$ so that for small angles, MRP converges towards Euler angles. In this work, a factor 1 is kept as per 6.

The time derivative of orientation is non linear (multiplicative) due to the nature of rotations.

$$\dot{q} = \frac{1}{2}\Omega(\omega)q \quad [7]$$

$$= \frac{1}{2}\begin{bmatrix}0\\ w\end{bmatrix} \otimes q \quad [8]$$

Where $\otimes$ denotes the quaternion product and Ω is defined as:

$$\Omega(\omega) = \begin{bmatrix} 0 & \omega_3 & -\omega_2 & \omega_1 \\ -\omega_3 & 0 & \omega_1 & \omega_2 \\ \omega_2 & -\omega_1 & 0 & \omega_3 \\ -\omega_1 & -\omega_2 & -\omega_3 & 0 \end{bmatrix} \quad [9]$$

The propagation equations can be derived when using MRP expression for attitude, as per [24]:

$$\dot{p} = f(p,\omega) = \frac{1}{2}\left[I\left(\frac{1-pp^\top}{2}\right) + [p_\times] + pp^\top\right]\omega \quad [10]$$

For the attitude, the system being non linear, the state vector is propagated by numerical schemes. A Runge-Kutta propagation model of order 4 is used.

4.3 Measurement model

Measurements from the image processing correspond to 2D keypoints directly extracted in the image. Each keypoint in the image is associated to its coordinates in the 3D satellite model. These measurements are integrated in a "tightly coupled" way since they are directly ingested in the Kalman filter.

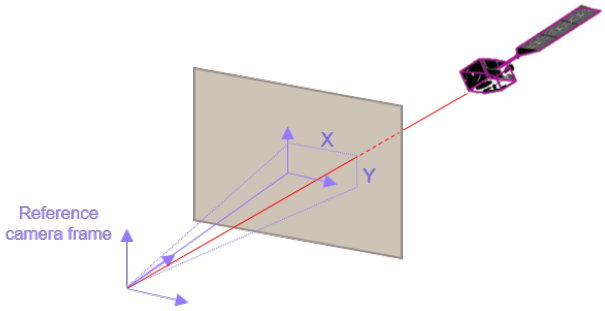


Fig. 8. A pinhole camera model is implement to project keypoints from client 3D model to image coordinates.

Image measurements are modeled as following the pin hole camera model (see figure 8) in the camera frame:

$$m_x = f_x\frac{x_c}{z_c} + U_x \text{ and } m_y = f_y\frac{y_c}{z_c} + U_y \quad [11]$$

where x_c, y_c and z_c are the keypoints coordinates in the camera frame, and f and U are camera parameters.

As a generic formulation, each 3D keypoint coordinates can be expressed as:

$$M_S = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = RM_C + t \quad [12]$$

where M_C corresponds to the keypoint coordinates in client object reference frame.

$$m_i = \frac{P_{cam}(RM_{Ci} + t)}{1_Z^T(RM_{Ci} + t)} \quad [13]$$

Where:

- m_i are the 2D positions of 3D points when projected to the image frame
- R is corresponds to the rotation matrix between object frame to camera frame
- t is the translation from object reference frame to camera frame
- M_{Ci} is any 3D point corresponding to the object 3D model
- 1_Z^T corresponds to the projection operator to Z camera axis
- P_{cam} is the camera projection matrix:

$$P_{cam} = \begin{bmatrix} f_x & 0 & U_x \\ 0 & f_y & U_y \end{bmatrix}$$

The previous measurement equations are expressed in the servicer camera frame and must be related to the respective attitudes and positions of the servicer and client. For any coordinates given in the client frame M_C, the quantity is computed in the servicer camera frame as M_S:

$$M_S = q_{C/S}^{-1} \otimes M_C \otimes q_{C/S} + t_{C/S}^S \quad [14]$$

with

$$q_{C/S} = q_{C/LOF} \otimes q_{LOF/I} \otimes q_{S/I}^{-1} \quad [15]$$

and

$$t_{C/S}^S = -q_{S/I} \otimes q_{LOF/I}^{-1} \otimes \begin{bmatrix} X_{S/C}^{LOF} \\ Y_{S/C}^{LOF} \\ Z_{S/C}^{LOF} \end{bmatrix} \otimes q_{LOF/I} \otimes q_{S/I}^{-1} \quad [16]$$

$q_{S/I}$ represents the inertial attitude of the servicer and is given by the on-board inertial filter. $q_{LOF/I} = \Omega_{orbital}$ is the rotation between LOF frame and inertial frame.

4.4 Retained architecture

Finally, the state vector representation is the following:

$$X_{pos} = \begin{bmatrix} X_{S/C}^{LOF} \\ Y_{S/C}^{LOF} \\ Z_{S/C}^{LOF} \\ \dot{X}_{S/C}^{LOF} \\ \dot{Y}_{S/C}^{LOF} \\ \dot{Z}_{S/C}^{LOF} \end{bmatrix} \quad X_{attitude} = \begin{bmatrix} p_{C/LOF}^{LOF} \\ \omega_{C/LOF}^{LOF} \end{bmatrix} \quad [17]$$

$$X = \begin{bmatrix} X_{pos} \\ X_{attitude} \end{bmatrix} \quad [18]$$

Where $X_{S/C}^{LOF}$ corresponds to the servicer position given in the LOF frame centered on the client. $p_{C/LOF}^{LOF}$ are the Modified Rodrigues Parameters representing the client orientation with respect to the Local Orbital Frame (LOF) and expressed in the LOF. $\omega_{C/LOF}^{LOF}$ is the rotational velocity of the client with respect to the inertial frame and expressed in the inertial frame. This representation is suitable for a kinematic filter. Inertial attitude information from servicer is easily integrated into the measurement model through equation 14.

At small client attitude angles and kinematics, the attitude propagation model is close to being linear. However, the measurement model is clearly non-linear and more difficult to being approximated by a linear function, especially at close distances. Therefore, a Unscented Kalman Filter (UKF) implementation is chosen to improve convergence and performance of the filter.

4.5 Outlier rejection

Inspired by [15, 16], keypoint outliers are identified and discarded at each filter udpate step based on Mahalanobis distance using estimated covariances and filter innovation.

Given the updated covariance at time k $\mathbf{S}_k$, the Mahalanobis distance of each measured keypoint i is computed as:

$$M_i = \sqrt{\mathbf{\Delta}_{i,k}^z \mathbf{S}_k^{-1} {\mathbf{\Delta}_{i,k}^z}^\top} \quad [19]$$

Where the innovation $\mathbf{\Delta}_{i,k}^z$ is defined for the measured keypoint i from the measurement function h and best state vector estimate at time step k $\tilde{x}_k$:

$$\mathbf{\Delta}_k^z = \mathbf{z} - h\left(\tilde{\mathbf{x}}_k\right) \quad [20]$$

Assuming gaussian distribution with means and covariances produced by the image processing algorithm, the Mahalanobis distances of keypoints should follow a χ law with 2 degrees of freedom. A threshold can be given over the distance corresponding to the probability that a keypoint be part of the assumed distribution. All keypoints above this distance are filtered out as outliers in the sense that the innovation measured does not match the covariance estimate from the image processing. As demonstrated in figure 9 the method allows to discard measurements in the keypoint error distribution tail.

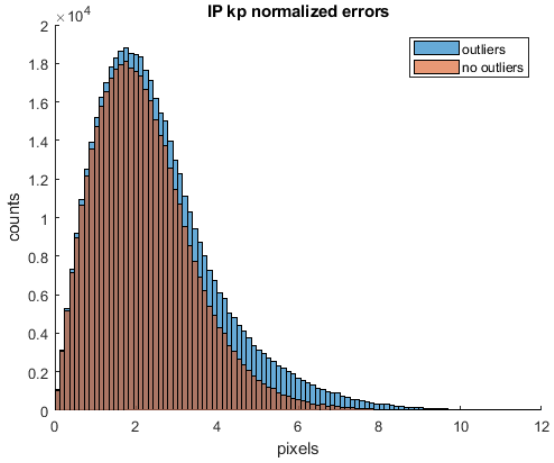


Fig. 9. Distribution of keypoint Mahalanobis residuals at the output of image processing. The outlier rejection method allows to reduce the distribution tail at the entrance of the navigation filter.

4.6 Convergence criteria

The outlier detection based on Mahalanobis distance can only be useful when estimated covariance from the filter is valuable. The initialization of the filter generally relies on poorly a priori initialized covariances. Therefore, the outlier rejection function is only applied after convergence of the UKF. Moreover a convergence status is also needed at AOCS level to decide to start the approach as soon as the filter outputs are stable.

The proposed criteria relies on the 2 following assertions:

1. The number of keypoints considered as outliers (based on the criteria presented above) is below a given threshold

2. The median innovation is below typical keypoint measurement errors as analyzed over the test image dataset

5. Performances

Performances are demonstrated over the dataset of 160 sequences presented in section 3.1.

The navigation chain performances (including image processing and navigation filter) are analyzed over the trajectory dataset with additional noises corresponding to on-board knowledge errors affecting the measurements and propagation models.

Parameter	Error type	Allocation
Servicer commanded thrust	Multiplicative gaussian random noise	$\sigma = 0.1\%$
Servicer attitude knowledge	Additive gaussian random noise	$\sigma = 10\mu rad$
Initial position knowledge	Additive gaussian random noise	$\sigma = 1\%$ of distance
Client attitude control residuals	Additive gaussian random noise	$\sigma = 1mrad$

To analyze the proper behaviour of the navigation filter, keypoint residuals are compared at the output of the Image Processing (IP) and at the output of the navigation filter (see 10). The navigation filter efficiently filters out keypoint outliers and improve residuals. The overall pipeline allows reaching subpixelic keypoint residuals thanks to regression heads.

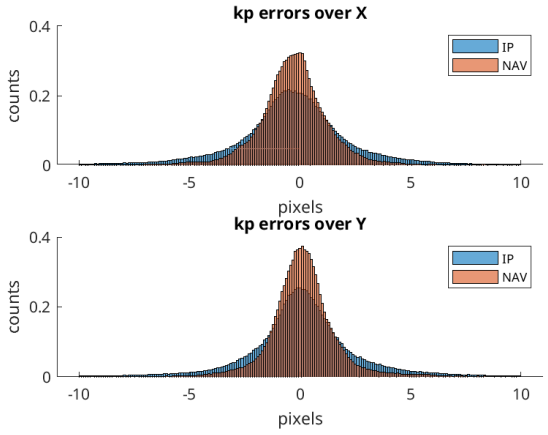
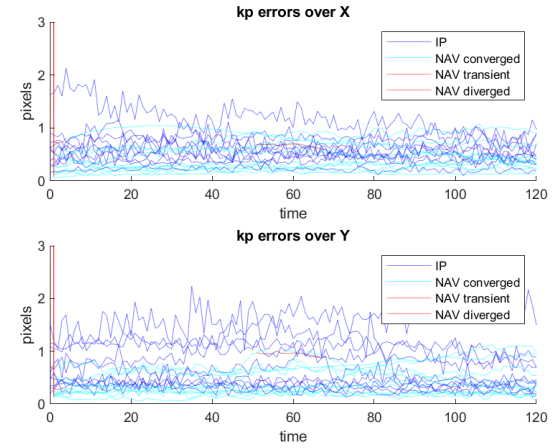


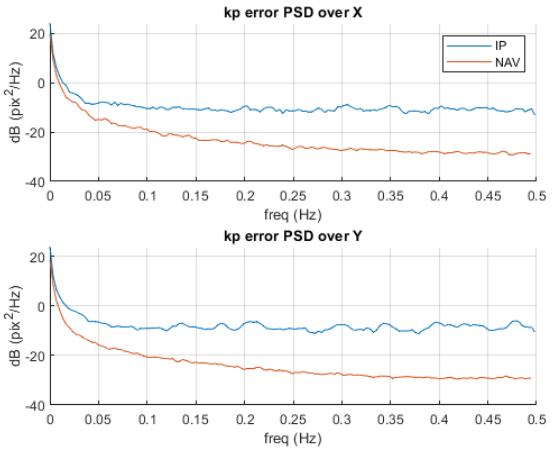
Fig. 10. Keypoint error distributions (at IP output and after navigation filter (given in full resolution image pixels)

The analysis of keypoint residual Power Spectral Density (PSD) clearly demonstrates a gain of almost 20dB at high frequency.

The pose restitution performances are compared against mission requirements (see figures 12a and 12b).



(a) Residuals over time

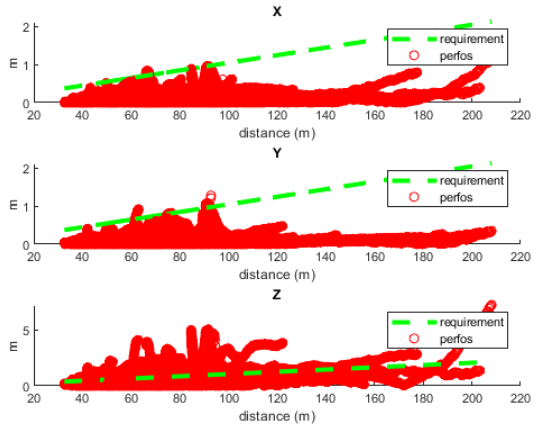


(b) Residual PSD

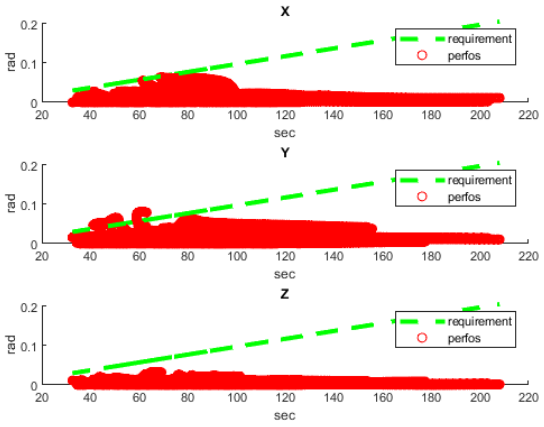
Fig. 11. Keypoint error residuals (at IP output and after navigation filter

The most challenging estimate remains the distance between servicer and client by the nature of image acquisition process. However, over the other axes, the requirements are reached.

Figure 13 illustrates that the probability of being outside relative position requirements in X and Y is below 1E-3 while more than 70% of the samples are within requirements in Z axis.



(a) Position residuals with respect to requirement (1%m of distance)



(b) Attitude residuals with respect to requirement (1/1000rad of distance)

Fig. 12. Relative pose residuals with respect to distance between client and servicer.

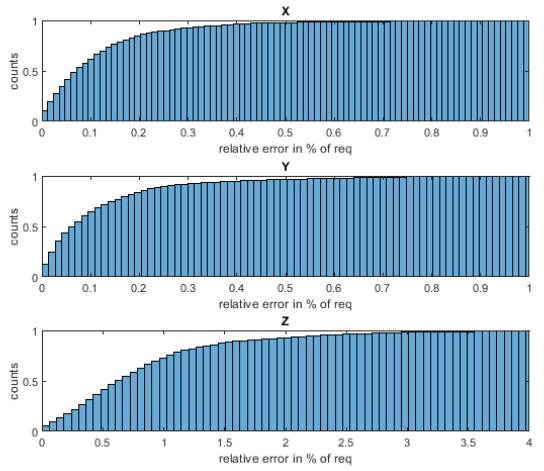


Fig. 13. Relative position error distribution with respect to requirements (outside requirements when above 1)

Regarding the number of processing steps required to reach the convergence criteria described in 4.6, performances are satisfactory with more than 90% of the sampled converged in less than 3 seconds (see figure 14). This again demonstrates the proper behaviour of the UKF filter and the efficiency of the initialization process.

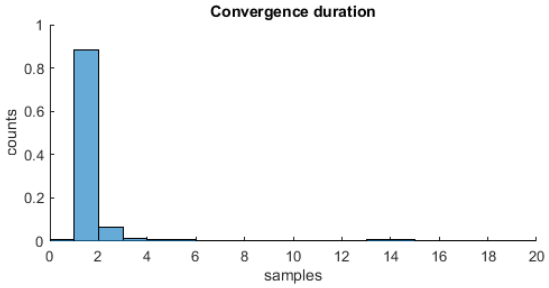


Fig. 14. Distribution of convergence duration of the navigation filter

6. Discussion

The work performed has demonstrated both:

- The relevance of the proposed image processing architecture that offers both very good precision and robustness.
- The adequacy of the UKF keypoint based navigation filter for cooperative scenarios.

The primary limitations are related to the accuracy of estimating the relative distance between the servicer and client satellites. Although a significant improvement of image processing outputs has been achieved with the use of a navigation filter, which enhances the observability of the range parameter, the performance still falls short of the required standards.

This issue could be addressed by either incorporating an additional range measurement sensor or optimizing the navigation architecture for better range estimation.

Regarding robustness with respect to domain gap between image simulations and real word data, a first analysis has been achieved in [18] demonstrating the robustness of such CNN architecture. However, a demonstration of the navigation chain robustness in real world conditions, typically on a test bench is still to be realized.

7. Conclusion

This papers introduces an end-to-end navigation architecture for vision-based cooperative rendezvous.

The image processing architecture is based on a light CNN backbone with specifically designed heads allowing both the regression of image keypoints and the estimation of error covariance. The navigation filter architecture proposed uses a UKF implementation offering proper performances in the presence of model non linearities. The proposed filter is specifically developed for cooperative rendezvous where the client satellite is still under control. Keypoint measurements from image processing are integrated as observables of the filter together with real time error estimates. Performances of the end-to-end architecture are demonstrated over a large set of synthetic trajectories offering the possibility to achieve detailed analysis of errors distribution and filter behaviour. Finally, specific developments have been proposed to solve for the initialization of the filter.

8. Acknowledgments

This work has been carried out by **IRT Saint-Exupéry** in the frame of **RAPTOR** project. This project has been lead and funded on behalf of **Thales Alenia Space**. The development of the image database and the validation of the navigation functions have been permitted thanks to the use of ©**SPICaM** as realistic image renderer tool.

References

[1] L. Pasqualetto Cassinis, R. Fonod, and E. Gill, "Review of the robustness and applicability of monocular pose estimation systems for relative navigation with an uncooperative spacecraft," en, *Progress in Aerospace Sciences*, vol. 110, p. 100 548, Oct. 2019, ISSN: 03760421. DOI: 10.1016/j.paerosci.2019.05.008.

[2] A. Comellini, F. Mave, V. Dubanchet, D. Casu, E. Zenou, and C. Espinosa, "Robust navigation solution for vision-based autonomous rendezvous," en, in *2021 IEEE Aerospace Conference (50100)*, Big Sky, MT, USA: IEEE, Mar. 2021, pp. 1–14, ISBN: 978-1-72817-436-5. DOI: 10.1109/AERO50100.2021.9438241. [Online]. Available: https://ieeexplore.ieee.org/document/9438241/.

[3] V. Capuano, K. Kim, A. Harvard, and S.-J. Chung, "Monocular-based pose determination of uncooperative space objects," en, *Acta Astronautica*, vol. 166, pp. 493–506, Jan. 2020, ISSN: 00945765. DOI: 10.1016/j.actaastro.2019.09.027.

[4] T. Hodan *et al.*, "Bop challenge 2023 on detection, segmentation and pose estimation of seen and unseen rigid objects," no. arXiv:2403.09799, Apr. 2024, arXiv:2403.09799 [cs]. DOI: 10.48550/arXiv.2403.09799. [Online]. Available: http://arxiv.org/abs/2403.09799.

[5] T. H. Park *et al.*, "Satellite pose estimation competition 2021: Results and analyses," en, *Acta Astronautica*, vol. 204, pp. 640–665, Mar. 2023, ISSN: 00945765. DOI: 10.1016/j.actaastro.2023.01.002.

[6] T. H. Park and S. D'Amico, "Robust multi-task learning and online refinement for spacecraft pose estimation across domain gap," *Advances in Space Research*, S0273117723002284, Mar. 2023, arXiv:2203.04275 [cs], ISSN: 02731177. DOI: 10.1016/j.asr.2023.03.036.

[7] Y. Bukschat and M. Vetter, "Efficientpose: An efficient, accurate and scalable end-to-end 6d multi object pose estimation approach," no. arXiv:2011.04307, Nov. 2020, arXiv:2011.04307 [cs]. DOI: 10.48550/arXiv.2011.04307. [Online]. Available: http://arxiv.org/abs/2011.04307.

[8] K. Park, T. Patten, and M. Vincze, "Pix2pose: Pixel-wise coordinate regression of objects for 6d pose estimation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 7668–7677.

[9] P. Castro and T.-K. Kim, "Crt-6d: Fast 6d object pose estimation with cascaded refinement transformers," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 5746–5755.

[10] Y. Su *et al.*, "Zebrapose: Coarse to fine surface encoding for 6dof object pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6738–6748.

[11] A. Lotti, D. Modenini, P. Tortora, M. Saponara, and M. A. Perino, "Deep learning for real time satellite pose estimation on low power edge tpu," no. arXiv:2204.03296, Jun. 2022, arXiv:2204.03296 [cs]. DOI: 10.48550/arXiv.2204.03296. [Online]. Available: http://arxiv.org/abs/2204.03296.

[12] Z. Wang, M. Chen, Y. Guo, Z. Li, and Q. Yu, "Bridging the domain gap in satellite pose estimation: A self-training approach based on geometrical constraints," no. arXiv:2212.12103, Dec. 2022, arXiv:2212.12103 [cs]. DOI: 10.48550/arXiv.2212.12103. [Online]. Available: http://arxiv.org/abs/2212.12103.

[13] J. I. B. Pérez-Villar, Á. García-Martín, and J. Bescós, "Spacecraft pose estimation based on unsupervised domain adaptation and on a 3d-guided loss combination," no. arXiv:2212.13415, Dec. 2022, arXiv:2212.13415 [cs]. DOI: 10.48550/arXiv.2212.13415. [Online]. Available: http://arxiv.org/abs/2212.13415.

[14] L. Pasqualetto Cassinis, R. Fonod, E. Gill, I. Ahrns, and J. Gil-Fernández, "Evaluation of tightly- and loosely-coupled approaches in cnn-based pose estimation systems for uncooperative spacecraft," en, *Acta Astronautica*, vol. 182, pp. 189–202, May 2021, ISSN: 00945765. DOI: 10.1016/j.actaastro.2021.01.035.

[15] L. P. Cassinis *et al.*, "Leveraging neural network uncertainty in adaptive unscented kalman filter for spacecraft pose estimation," en, *Advances in Space Research*, vol. 71, no. 12, pp. 5061–5082, Jun. 2023, ISSN: 02731177. DOI: 10.1016/j.asr.2023.02.021.

[16] T. H. Park and S. D'Amico, "Adaptive neural network-based unscented kalman filter for robust pose tracking of noncooperative spacecraft," *Journal of Guidance, Control, and Dynamics*, vol. 46, no. 9, pp. 1671–1688, Sep. 2023, arXiv:2206.03796 [cs, eess], ISSN: 0731-5090, 1533-3884. DOI: 10.2514/1.G007387.

[17] *Runge kutta methods*, https://en.wikipedia.org/wiki/Perspective-n-Point, Accessed: 2024-09-18.

[18] F. Gallet, C. Marabotto, and T. Chambon, "Exploring ai-based satellite pose estimation: From novel synthetic dataset to in-depth performance evaluation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 6770–6778.

[19] A. Howard *et al.*, *Searching for mobilenetv3*, 2019. arXiv: 1905.02244 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1905.02244.

[20] M. Tan, R. Pang, and Q. V. Le, *Efficientdet: Scalable and efficient object detection*, 2020. arXiv: 1911.09070 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1911.09070.

[21] D. A. Nix and A. S. Weigend, "Estimating the mean and variance of the target probability distribution," in *Proceedings of 1994 ieee international conference on neural networks (ICNN'94)*, IEEE, vol. 1, 1994, pp. 55–60.

[22] W. Fehse, *Automated rendezvous and docking of spacecraft* (Cambridge aerospace series), en. Cambridge: Cambridge University Press, 2003, ISBN: 978-0-521-82492-7.

[23] B. O. Lange and R. G. Smith, "The application of floquet theory to the computation of small orbital perturbations over long time intervals using the tschauner-hempel equations," Tech. Rep., 1965.

[24] H. Schaub and J. L. Junkins, "Stereographic orientation parameters for attitude dynamics: A generalization of the rodrigues parameters," en,

[25] J. L. Crassidis and F. L. Markley, "Unscented filtering for spacecraft attitude estimation," en,