VISION-BASED ALGORITHM AND ROBUST FILTERING FOR STATE ESTIMATION OF AN UNCOOPERATIVE OBJECT IN SPACE

Vincenzo Pesce[1], **Luca Losi**[1], **and Michèle Lavagna**[1]

[1]*Politecnico di Milano, Aerospace Science and Technology Department, via La Masa, 34, 20156 Milano, Italy, vincenzo.pesce@polimi.it, luca.losi@polimi.it, michelle.lavagna@polimi.it*

ABSTRACT

A Vision Based algorithm relying on a mono-camera as main sensor has been developed at PoliMi DAER to be used as main measurement tool to feed the Navigation filter for relative pose estimation of uncooperative objects. The Vision Based algorithm works combining a Perspective-*n*-Point problem (P*n*P) along with Bundle Adjustment (BA). This is an optimization technique widely diffused in Computer Vision, that has been applied on a built offline sparse map and on the 2D features extracted from the images coming from the camera to optimize the retrieved relative position and attitude, which are finally given in input to the navigation filter. A decoupled model is used for the relative translational and rotational dynamics. A first numerical validation of both the Vision based routine and the filter is presented by using, as input, measurements obtained from synthetic images. The obtained results are finally analyzed and critically discussed.

Key words: Visual tracking; Filtering; State estimation; Uncoperative objects.

1. INTRODUCTION

Nowadays all the major space agencies and industries are undertaking the enhancement of the automation capabilities of their spacecraft. Autonomous operations are extremely advantageous for space missions, with a wide range of possible applications. For example, autonomous relative navigation around unknown and uncooperative objects is particularly appealing. Precise pose and motion estimation of an uncooperative object, such as a Resident Space Object (RSO) has a potential utilization in the domain of space debris removal and on-orbit servicing. In the last decade, several different works have tried to solve this problem. This paper wants to investigate the potentiality of an hybrid approach, combining vision-based algorithm for pose estimation with robust filtering methods. Vision Based systems represent nowadays a promising tool to obtain satellite relative trajectory estimation within an optimal level of accuracy. Being this topic first born in the Computer Vision field, an increasing migration and development of these techniques to space applications is currently occurring, with specifically developed algorithms and hardware (e.g. more powerful and multicore CPUs) borned in recent years. Within this context, a Vision Based algorithm relying on a mono-camera as main sensor has been chosen and developed at PoliMi DAER along with a Navigation filter for relative pose estimation of uncooperative objects. The Vision Based algorithm is built to be used as main measurement tool to feed the Navigation filter.

This paper is organized as follows: in the first two parts the Vision Based tracking and the Navigation filter are presented and described with each of their constituent blocks; in the third part a first numerical validation of both the algorithm is presented by using, as input, measurements obtained from synthetic images. To validate and assess the software accuracy, synthetically generated trajectories, along with a point cloud model simulating a target object, and correspondent image projections have been firstly used. To conlude, the obtained results are finally analyzed and critically discussed.

2. SYSTEM OVERVIEW

The Vision Based algorithm implements a Visual Odometry like routine [1][2] and works detecting the target object features from the incoming images, given by the mono-camera; these are then matched to an already available on-board map (constituted by a mesh of 3D points, each one correlated to a descriptor) and, in this way, a set of 3D to 2D correspondences is built. From the set of correspondences the so called Perspective-*n*-Point problem (P*n*P) is built and solved within a RANSAC routine in order to delete incoming outliers (wrong match between target image and on-board map) and obtain a first estimate of the relative pose; relative position of the camera (and hence of the target-chaser) which is obtained in terms of rototranslation. Bundle Adjustment (BA), an optimization technique widely diffused in Computer Vision [3], is then applied on the map and on the 2D features (which constitute the measurements) in order to optimize the obtained pose, which is finally given in input to the navigation filter. The on-board 3D sparse map of the target object is built on-ground with a dedicated algorithm, by correlations of a 3D model of the target uncooperative object with descriptors extracted from multiple images of it. This first estimate of the pose (relative position and attitude) of the observed uncooperative object is

fed to the navigation filter. A decoupled model is used for the relative translational and rotational dynamics. A linear model, including eccentricity, is implemented and used as model of the dynamics inside the filter. Linear models allow for the exploitation of robust Kalman filter techniques. These methods result to be more reliable and robust to uncertainties. For the rotational dynamics, the non-linear equations constrain to use a non-linear filter such as the Extended Kalman Filter. However, the use of an alternative formulation, based on Lie-Groups, is here investigated. This formulation should allow the preservation of the natural symmetry of the physical system without any ambiguity or singularity.
Fig.1 shows the global structure of the relative navigation system. Vision Based tracking is represented with its main components: Features detection and matching, motion estimation and optimization. Pose retrieved from this block is then given in input to the filter, which manages separately translations and rotations to give the final relative state estimate (including velocities) of the target-chaser spacecraft.

3. VISION BASED TRACKING

In this section all the building blocks constituting the Vision Based tracking algorithm are described along with the steps that are performed with every frame of the camera.

3.1. Features detection

Features are salient points that can be localized on each image coming from the camera, and different algorithms for their detection exists [4][5][6][7]. Features detection is a fundamental step in the routine of the Vision Based tracking algorithm because it allows to build the correspondences with the available map that are then used to retrieve the relative pose of the target. The more robust are the features, the more accurate will be the matching with the *3D*-model resulting in a more accurate pose estimation.
A general in-orbit relative chaser-target spacecraft trajectory implies modest scale variations between different frames along with rotations; moreover, different light conditions will be met moving along the orbit. Features extracted, to be robust, shall be invariant to each of these parameters. Moreover, computational cost has to be considered and has to be as low as possible for real time application. ORB detector [7] has been selected for implementation considering these requirements. ORB is a feature detector and fast binary descriptor based on BRIEF[8] and FAST[6] whose features are extremely fast to compute, have good invariance to viewpoint, scale and are resilient to different light conditions. At the moment ORB detector has been implemented to work building an $8-$Level grey scale pyramid with a scale factor of 1.2 for each incoming image. An upper bound of 300 features along with their descriptors is set to be extracted to maintain low the computational cost. Fig.2 shows an example of extracted features on a generic image of Suomi-NPP satellite. Green markes show features location along with their pyramid level and orientation. Substitution of the actual routine with background subtraction methods will be carefully evaluated in the future both in terms of performances of the tracking algorithm and computational costs in order to directly separate spacecraft image from the background and extract an higher number of features on it avoiding unnecessary zones.

3.2. Features matching with 3D-model

The matching step in the algorithm consists in building a set of correspondences between the detected observed *2D* features in the current frame and the points of the available target spacecraft *3D*-model. Matching is performed exploiting *Hamming distance* as linearity measure between ORB descriptors extracted in the image and descriptors of the *3D*-model. Hamming distance can be computed very efficiently between corresponding binary descriptor strings and makes the process very fast. For each descriptor from the frame the matcher finds the two closest descriptors (in order of score) in the *3D*-model by trying one by one. Once the set of matches is obtained, a ratio test according to [4] is applied: the distance of the closest descriptor to that of the second-closest descriptor is compared; correct matches need to have the closest descriptor significantly closer than the closest incorrect match to achieve reliable matching. Only matches with a distance ratio lower than a fixed threshold are therefore retained. Ratios between 0.7 and 0.8 actually ensures the better performances. This is an heuristic way to immediately discard clear outliers and retain only good matches.

3.3. Motion estimation with P*n*P

The *2D* features from frame to *3D*-model correspondences obtained from the features matching step are used to solve the "Perspective-*n*-Point" problem in this sub-block of the routine. The P*n*P problem is namely the problem of estimating the pose of a calibrated camera given a set of *n 3D* points in the world reference and their corresponding *2D* projections on the image.
The *EPnP* algorithm [9] is exploited by the Vision Based tracking, which implements an efficient solution to the problem, it is non-iterative and applicable for both planar and non-planar *3D* clouds configurations. The algorithm works within a *RANSAC* [10] routine in order to be robust to the presence of outliers and gives a first estimation of the relative position and orientation of the chaser camera with respect to the target.

3.4. Motion only Bundle Adjustment

Motion estimation by itself does not give a sufficient accuracy in trajectory reconstruction. Uncertainty in extracted features location due to noise, presence of outliers between the *2D-3D* correspondences and uncertainty in the built *3D*-model all result in a pose estimation from the *EPnP* algorithm which drifts from the true trajectory. Bundle Adjustment (BA) [3]is therefore implemented to counteract these effects and to correct the trajectory estimation. BA technique is designed to

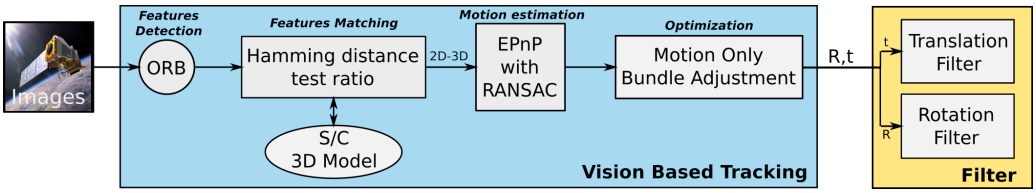

Figure 1: Relative Navigation System architecture. Vision Based tracking and Filter are shown with their component.

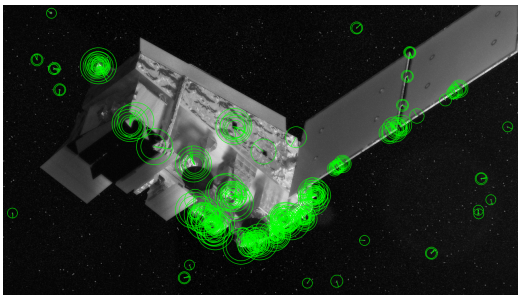

Figure 2: Example image with ORB features extracted. Green markers indicate pyramid level and orientation of the feature.

optimize camera pose minimizing a function that is the error in reprojecting *3D* points whose *2D* coordinates are known on the images and taking as constraint the scene geometry.
g2o, a library specifically developed for pose graph optimization, is exploited as tool for the BA implementation. Pose graph is a way of formulating Simultaneous Localizantion and Mapping (SLAM) problem in which graph are used whose nodes represent map points and poses of the robot in time and whose edges represent constraints between the poses (measurements). Once such graph is constructed, BA is applied to find a configuration of the nodes which is maximally consistent with the measurements
This approach has been implemented for the Vision Based tracking in the so called *"Motion only BA"* version: At the end of each motion estimation step, retrieved camera pose and matched *3D*-model points seen from the current frame are set as vertices of the graph, while observed features representing the measurements define the edges. The obtained graph is optimized keeping fixed the map points and improving the camera pose estimation only, which is then given in input to the navigation filter.

4. FILTERING

Once the images are processed and the measurements are available, it is possible to perform a filtering procedure. In this work, decoupled filters are used, one for the translation and another for the relative attitude estimation.

4.1. Translation filter

For the translational part of the filter, an H-∞ Filter has been used. This kind of filter guarantees robustness minimizing the worst case estimation error [11]. Let's consider a linear time-invariant system:

$$\begin{aligned} x_{k+1} &= F_k\, x_k + B\, w_k \\ y_k &= H_k\, x_k + D\, v_k \end{aligned} \tag{1}$$

with x_k the state vector, w_k and the v_k the process and measurement noise respectively and y_k the measurement output. Similarly to the Kalman Filter, the correction equation can be defined as:

$$\hat{x}_{k+1} = F_k\, \hat{x}_k + F_k\, K_k\, (y_k - H_k\, \hat{x}_k) \tag{2}$$

However, K has not the same expression of the Kalman Filter. In particular we want to find K such that $\|T_{ew}\|_\infty < \dfrac{1}{\theta}$, where T_{ew} represents the difference between the predicted and real state and θ is a tuning parameter. To find the value of K it is possible to explicit a Riccati-like equation:

$$P_{K+1} = F_k P_k [I - \theta P_k + {H_k}^T {R_k}^{-1} H_k P_k]^{-1} F^T + Q_k \tag{3}$$

And the expression of the gain is:

$$K_k = P_k [I - \theta P_k + {H_k}^T {R_k}^{-1} H_k P_k]^{-1} {H_k}^T {R_k}^{-1} \tag{4}$$

A linearized model for the dynamics has been selected. In particular, in this work, the authors used the formulation by Yamanaka and Ankersen [12]. Their derivation of the state transition matrix can be very advantageous for the implementation of filtering techniques and control systems. In fact, an expression for the F matrix of eq.1 is directly derived. The details of the implementation are not presented here and they can be found in [12]. On the other side, the measurement model is easily derived since the relative position between the two centers of mass is directly the state. For this reason, the H matrix in eq.1 is equal to identity.

4.2. Rotation filter

For the rotation part, a minimum energy filter on the Lie group is implemented. Minimum energy filtering was firstly introduced by Mortensen [13]. Recently, it has

been specialized to attitude estimation on the Special Orthogonal Group SO(3) by Zamani et al. [14]. Our implementation is derived from the one presented by Zamani [15], however a slightly different formulation is hereby proposed. In fact, a simplification can be introduced without considering the real dynamics of the system. The main advantage of this approach is that it is not necessary to know the exact value of the inertia matrix of the target and the filter can be expressed directly in the camera reference frame. In particular:

$$\begin{aligned} \dot{R} &= R(\omega(t)_\times) \\ \dot{\omega} &= 0 + B\delta \end{aligned} \tag{5}$$

This lead to the following filter formulation:

$$\begin{aligned} \hat{R}(t_0) &= \hat{R}_0, \;\; \hat{\omega}(t_0) = \hat{\omega}_0, \;\; K(t_0) = K_0, \\ \dot{\hat{R}} &= \hat{R}((\hat{\omega}(t) + K_{11}\, r^R + K_{12}\, r^\omega)_\times, \\ \dot{\hat{\omega}} &= K_{21}\, r^R + K_{22}\, r^\omega \\ \dot{K}(t) &= -\alpha\, K + AK + KA^T - KEK + GQ^{-1}G^T \\ &\quad - WK - KW^T \end{aligned} \tag{6}$$

with

$$\begin{aligned} r_t &= \begin{bmatrix} r^R \\ r^\omega \end{bmatrix} = \begin{bmatrix} -(u_1)(\hat{r}_1 \times r_1) - (u_2)(\hat{r}_2 \times r_2) \\ 0 \end{bmatrix} \\ \hat{r}_i &= \hat{R}^T \overline{r}_i \qquad r_i = R^T \overline{r}_i + d_i\, \epsilon \\ A &= \begin{bmatrix} -\hat{\omega}_\times & I \\ 0 & 0 \end{bmatrix} \\ E &= \begin{bmatrix} \sum_{i=1}^{2} u_i((\hat{r}_i)_\times (r_i)_\times + (r_i)_\times (\hat{r}_i)_\times)/2 & 0 \\ 0 & 0 \end{bmatrix} \\ GQ^{-1}G^T &= \begin{bmatrix} 0 & 0 \\ 0 & GQ^{-1}G^T \end{bmatrix} \\ W &= \begin{bmatrix} \frac{1}{2}(K_{11}\, r^R + K_{12}\, r^\omega)_\times & 0 \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{7}$$

5. SIMULATION SCENARIO

In order to validate and asses the performance of both the Vision Based tracking and the filter, and of the whole navigation routine together, a preliminary test campaign has been performed on a simulated scenario. A MEO orbit has been selected for the chaser spacecraft with an eccentricity of 0.17 and semi-major axis of $8790\,\mathrm{km}$. The relative reference dynamics is simulated directly integrating the nonlinear equations for the relative motion without considering any orbital perturbation. The assumed initial conditions are:

$$\begin{aligned} \rho_0 &= [50, 0, 0] \qquad m; \\ \dot{\rho}_0 &= [0, -0.1, 0] \quad m/s; \end{aligned} \tag{8}$$

expressed in the local-vertical, local-horizontal (LVLH) reference frame fixed to the chaser spacecraft center of mass, with $\hat{x}$ directed radially outward, $\hat{z}$ normal to the chaser's orbital plane, and $\hat{y}$ completing the triad. This initial conditions have been chosen to have an in-plane elliptical motion of one spacecraft with respect to the other. This trajectory can be representative of a monitoring or close approach phase. It has to be underlined that the non perfect absolute state determination of the chaser spacecraft is taken into account. In particular, the true anomaly, input of the relative dynamics model, is obtained from an absolute state corrupted by noise. The noise associated to the absolute position is modeled as a normal distribution with $0\,\mathrm{mean}$ and a standard deviation of $10 \times 10^{-2}\,\mathrm{km}$ and the noise of the velocity is assumed to be a normal distribution with $0\,\mathrm{mean}$ and a standard deviation of $10 \times 10^{-4}\,\mathrm{km\,s^{-1}}$. For the relative dynamics, a torque-free tumbling motion has been imposed to the simulated target spacecraft. The motion has been simulated using the classical Euler equation for rigid body, imposing the following initial conditions: $\omega_0 = [1, 0.1, 0.3] deg/s$. This values have been used to reproduce a generic tumbling motion with significant angular velocity. This can be representative of a tumbling space debris. Target spacecraft has been modelled with a simple shape taking as reference the NASA Suomi-NPP satellite. A set of 38 uniform distribute points has been considered and constitutes the *3D*-model used by the Vision Based tracking. The set of images to be provided to the Navigation algorithm have been instead directly obtained by projecting the *3D*-model points onto the image plane. To take into account for the uncertainty in features detection that arise due to the presence of noise for an application on real images, a uniform distributed noise between 0 and 2 pixels has been added to the *2D* features location. This error, when at maximum level, already generates some outliers that are consequently inserted in the set of *2D-3D* correspondences. Fig.3 shows the true *3D* built model of the spacecraft along with the selected features, both with zero noise. Fig.4 shows instead a short sequence of the generated synthetic images with the correspondent features, from which it is possible to observe a portion of the spacecraft tumbling motion.

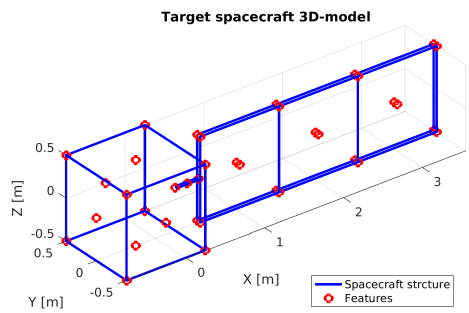

Figure 3: *3D* spacecraft model used for the simulations along with considered features.

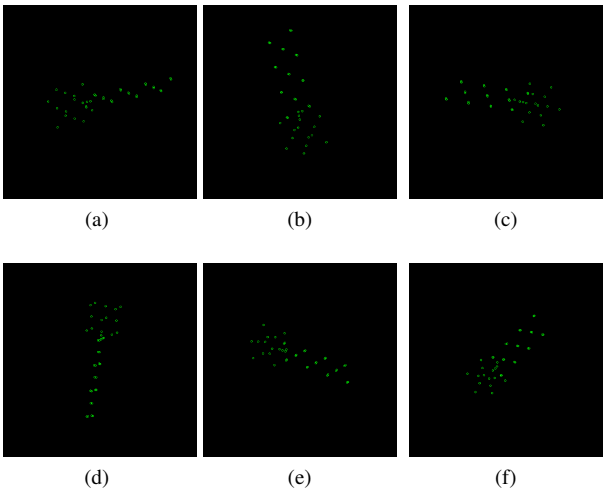

Figure 4: Simulated target spacecraft images with features.

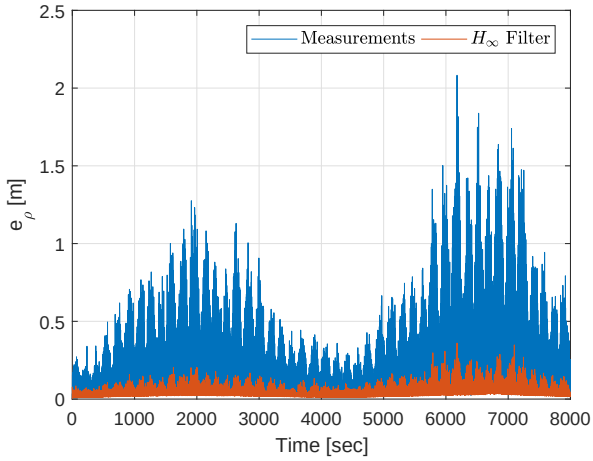

Figure 5: Relative Position Error

6. RESULTS

In this section, the results of the pose determination algorithm and of the overall relative state determination are presented. The estimation errors are computed as follows. The position error is computed as:

$$e_\rho = \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (z_i - \hat{z}_i)^2} \tag{9}$$

where $\hat{x}, \hat{y}, \hat{z}$ are the position components estimates. Similarly, the velocity error is:

$$e_{\dot{\rho}} = \sqrt{(\dot{x}_i - \hat{\dot{x}}_i)^2 + (\dot{y}_i - \hat{\dot{y}}_i)^2 + (\dot{z}_i - \hat{\dot{z}}_i)^2} \tag{10}$$

The relative attitude error is computed as:

$$e_R = acos\left(1 - \frac{tr(I - {R_i}^T \hat{R}_i)}{2}\right) \tag{11}$$

with $\hat{R}$ being the estimated rotation matrix.

The results of the translational H-∞ Fliter are presented in this paragraph along with the pose determination results. In particular, the estimation errors corresponding to the 2 pixels noise case with measurement frequency of 10 Hz are considered. Fig. 5 shows the estimation error of the position. Fig. 6 instead, shows the error of the corresponding relative translational velocity estimates for the H-∞ Filter.

The advantage of having a filter downstream of the image processing is evident. The overall position error is strongly reduced and it always stays beyond $0.4\,\mathrm{m}$. Moreover, very good results are obtained for the estimation of the relative translational velocity.

Similarly to what has been done for the translational filter, the results of the rotational filter are be presented.

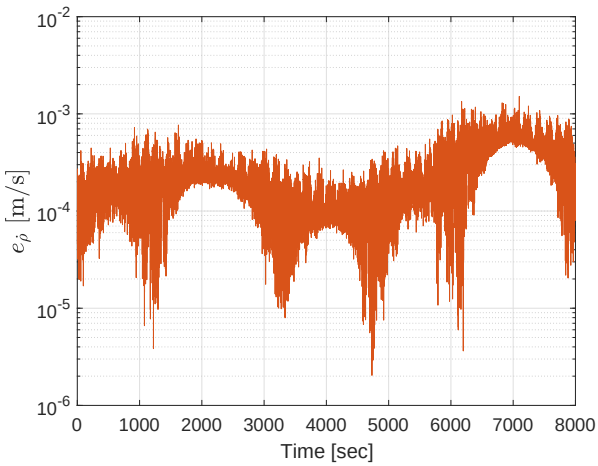

Figure 6: Relative Velocity Error

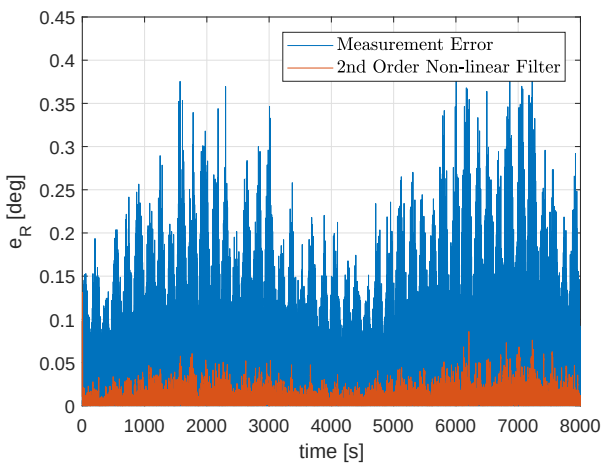

Figure 7: Relative Attitude Error

Fig. 7 show the rotation filtering results. Once again the filter has improved quite significantly the measurments from the Vision Based tracking, The overall rotation error is always beyond 0.1 deg.

7. CONCLUSION

This paper investigates the possibility of combining image processing techniques with robust filtering algorithms for relative state estimation of uncooperative space objects. The presented results show how the implemented solution allows to obtain a precise estimation of the relative position, velocity and attitude. The performance of the image processing algorithms have been preliminary assessed. The obtained results are very promising but a further analysis with real images is necessary. An extensive analysis is currently ongoing to test both image processing and filters with different scenarios. The presence of filters for the translational and rotational dynamics is a clear advantage for the relative state estimation. A more extensive analysis is also needed for the filtering part to assess its robustness and efficiency over different simulations conditions.
One of the next development steps will be the validation of the proposed algorithms with the experimental facility currently under setup at Politecnico di Milano Department of Aerospace Science and Technologies. The facility comprises a 7-DOF Mitshubishi PA-10 robotic arm which will be used to replicate the relative motion between a chaser and an uncooperative target. On the robotic arm end-effector, a 3D-printed satellite mock-up will be installed, with a fixed camera acquiring its images during motion. This setup will allow the validation of both image processing and pose determination algorithms under more realistic conditions up to TRL4.

REFERENCES

1. D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *Robotics & Automation Magazine, IEEE*, vol. 18, no. 4, pp. 80–92, 2011.

2. F. Fraundorfer and D. Scaramuzza, "Visual odometry: Part ii: Matching, robustness, optimization, and applications," *Robotics & Automation Magazine, IEEE*, vol. 19, no. 2, pp. 78–90, 2012.

3. B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis," in *Vision algorithms: theory and practice*. Springer, 1999, pp. 298–372.

4. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

5. H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer vision–ECCV 2006*. Springer, 2006, pp. 404–417.

6. E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Computer Vision–ECCV 2006*. Springer, 2006, pp. 430–443.

7. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2564–2571.

8. M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," *Computer Vision–ECCV 2010*, pp. 778–792, 2010.

9. V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o(n) solution to the pnp problem," *International journal of computer vision*, vol. 81, no. 2, pp. 155–166, 2009.

10. M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

11. D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.

12. K. Yamanaka and F. Ankersen, "New state transition matrix for relative motion on an arbitrary elliptical orbit," *Journal of guidance, control, and dynamics*, vol. 25, no. 1, pp. 60–66, 2002.

13. R. Mortensen, "Maximum-likelihood recursive nonlinear filtering," *Journal of Optimization Theory and Applications*, vol. 2, no. 6, pp. 386–394, 1968.

14. M. Zamani, J. Trumpf, and R. Mahony, "Minimum-energy filtering for attitude estimation," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2917–2921, 2013.

15. A. Saccon, J. Trumpf, R. Mahony, and A. P. Aguiar, "Second-order-optimal minimum-energy filters on lie groups," *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 2906–2919, 2016.