ON-ORBIT RELATIVE NAVIGATION NEAR A KNOWN TARGET USING MONOCULAR VISION AND CONVOLUTIONAL NEURAL NETWORKS FOR POSE ESTIMATION

i-SAIRAS, Virtual Conference, 19–23 October 2020

Arunkumar Rathinam[1] and Yang Gao[2]

[1]STAR LAB, Surrey Space Centre, University of Surrey, Guildford, UK. Email: a.rathinam@surrey.ac.uk
[2]STAR LAB, Surrey Space Centre, University of Surrey, Guildford, UK. Email: yang.gao@surrey.ac.uk

ABSTRACT

In recent years, there is an increasing demand for orbital robotic missions for various reasons such as life-extension of functional satellites, reuse the unique orbital slots and to reduce the risk of orbital collision. In such robotic missions, the satellite's autonomous navigation capability is a critical component that enables it to perform relative navigation, inspection, and repair with minimal human-in-loop intervention. Pose estimation is an important task within autonomous GNC for spacecraft in orbit. There have been recent, new development of deep learning based pose estimation algorithms in order to meet growing demands of autonomous orbital applications. This paper presents a new keypoint-based framework using Convolutional Neural Network models for pose estimation of known non-cooperative targets in orbit, which is thoroughly compared to existing state-of-the-art algorithms also based on deep learning. Within the proposed pose estimation pipeline, a ResNet-based architecture used for object detection, a Scale-Aware High-Resolution Network (HigherHRNet) used for keypoint regression and PnP-RANSAC for computing the pose. The framework is benchmarked with the SPEED dataset as well as the Soyuz dataset from STAR LAB Orbital Visual Simulator and the results were presented.

1 INTRODUCTION

In the last decade, there is a growing interest in orbital robotic missions to autonomously carry out On-Orbit Servicing (OOS) and Active Debris Removal (ADR). To enable a successful orbital robotic mission, the modern Guidance, Navigation, and Control (GNC) solutions must support more autonomous functions like relative navigation, rendezvous, and other manipulations such as capturing the target or debris. OOS and ADR are considered key capabilities for spaceflight in this century and multiple technology demonstration missions including PROBA-3 [1] by ESA, PRISMA [2] by OHB Sweden. Recently, the first commercial OOS of a geostationary satellite (IntelSat-901) carried out by Space Logistics using the MEV-1 (Mission Extension Vehicle) satellite platform. MEV-1 docked with the IntelSat satellite and re-positioned it to the designated spot and continues to provide in-orbit station-keeping services [3].

Both OOS and ADR mission operations involve orbital rendezvous with the target before performing relative navigation at the close proximity. Multiple sensor options are available to perform relative navigation, however, monocular cameras are widely considered because of the lower hardware complexity, cost, weight and power consumption. As with every sensor, there are limitations associated with the monocular camera such as its inability to provide depth measurements, sensitivity to adverse illumination conditions. During close proximity operations, the GNC system needs an estimate of 6 Degree-of-Freedom (DOF) pose of the target, i.e., the relative position and attitude, that represents a piece of key information for the navigation system.

This work discusses the deep-learning framework for spacecraft pose estimation using keypoints for relative navigation. The framework presented is within the family of keypoint-based methods and its performance represents the state-of-the-art solutions in pose estimation for known non-cooperative targets in orbit. This paper further shows qualitative comparisons and analysis of two representative approaches or families of methods for deep learning based pose estimation, namely keypoint approach and non-keypoint approach (also known as direct approach that performs regression/classification directly on pose data). The results have been obtained using two datasets for testing and validation: ESA-Stanford's benchmark dataset, Spacecraft PosE Estimation Dataset (SPEED) based on PRISMA mission [4] and the photo-realistic Soyuz dataset generated by STAR LAB's Orbit Visual Simulator (OrViS) [5]."

2 BACKGROUND

Estimating the camera pose, i.e., the position and orientation, from a single image is a fundamental computer vision problem. The camera pose represents critical information to many robotic applications such as

localization and navigation. During the relative pose estimation process, the algorithm predicts the rigid-body transformation from the object's coordinate system to the camera coordinate system.

The major approaches that are used today for the pose estimation and tracking are the fiducial-, model- and non-model-based approaches. The fiducial detection and tracking utilize the known markers on the tracked objects and identification of the detected fiducials allow to match 2D image features with their calibrated 3D features. Fiducials are suitable for repetitive close-range applications, but it is impractical to use in all cases. Model-based techniques rely on the prior knowledge of the target object and the target's pose can be estimated via feature-detection and matching, following non-linear pose refinement. Whereas, Non-Model-based techniques do not assume prior knowledge of the target object's geometry, texture or other visual identifications. These methods solve for the optimal camera pose or motion through recovering identical features in the images and under epi-polar or motion field constraints.

One of the main challenge faced by the feature point extraction methods is a sharp change in shadows, appearance change due to rotation/tumbling motion, the low Signal-To-Noise Ratio (SNR) and the high contrast which characterize the space images. Many of the algorithms may have difficulties with image-to-model feature correlation, feature persistence for tracking over more than a few frames, foreground-background segmentation, and detection and correction of correlation and tracking errors.

3 LITERATURE REVIEW

With recent advancements in deep learning and the popularity of the ESA's Spacecraft Pose Estimation Challenge enabled new developments with the state-of-the-art performance in the visual pose estimation algorithms. Several research works and their results are published based on this competition including Spacecraft Pose Network (SPN) [4], Pose Estimation with Deep Landmark Regression [6], Pose Estimation with soft classification [5], segmentation driven approach [7].

SPN [4] used a combination of classification and regression approaches to computing the relative pose. It predicts the bounding box of the satellite in the image with an object detection network and the bounded sub-image is processed through a sub-network to perform classification on the 3D pose. During the classification, the SO3 rotation group is discretized into m uniformly distributed fine-bins representing the base rotations and the network retrieves the n-most relevant rotations from the feature map of the detected object. The translation of the pose is estimated via the constraints from the bounding box dimensions to fit the entire object with the predicted rotation. Chen et.al [6] presented a keypoint-based approach to estimate the pose of the satellite. They regress the bounding box around the satellite using an object detection CNN and crop the image. The cropped image is then fed into the keypoint regression CNN to obtain the 2D locations of the landmarks. Finally, the 2D-3D landmark correspondences and non-linear optimization used to compute the pose estimates. Gerard [7] presented a segmentation-driven approach where the segmentation stream used to identify the object region and regression stream to predict 2D keypoint position. Finally, the iterative PnP with a RANSAC-based version of the EPnP is used to compute the pose. One of the limitations is that the wrong prediction of 2D keypoint will lead to inaccurate PnP pose estimate. Proenca and Gao [5] proposed a hybrid approach which involves a regression branch for location estimation and a probabilistic soft-classification framework to predict the orientations. Dhamani et. al [8] presented a CNN-based approach to estimate the relative bearing (azimuth and elevation). The algorithm was developed for and deployed on the Seeker-1 mission, a CubeSat class technology demonstrator mission, intended to provide relative bearing estimates of the non-cooperative Cygnus vehicle from ranges of 5 to 40 meters in real-time ($>$1 Hz). Harvard et.al [9] proposed an architecture that uses existing keypoint localization algorithms to identify robust keypoints and then train a CNN on this limited set of keypoints (with feature descriptor components) to create specialized descriptors. For each landmark, a visibility map generated through ray tracing. PnP-RANSAC was used to estimate the pose and non-linear filter to track the pose.

4 POSE ESTIMATION FRAMEWORK

The pose-estimation framework used in this work is similar to the approach presented in [6]. The framework utilizes object detection, followed by the keypoints estimation, finally the PnP-RANSAC to estimate the pose of the target object. Two models from different datasets were tested in this work, and they are the TANGO spacecraft in the SPEED dataset and the Soyuz spacecraft from the STAR LAB's OrViS (previously named as URSO in [5])

4.1 Object detection

To detect the bounding box of the image, a ResNet-based Faster R-CNN used as the backend. ResNet-50 [10] with the pre-trained weights from the COCO dataset used to detect the bounding box of the object. To train the ResNet algorithm for bounding box detection the ground truth coordinates were obtained from the keypoint coordinates with some relaxations to the lengths between the minimum and maximum pixel coordinates.

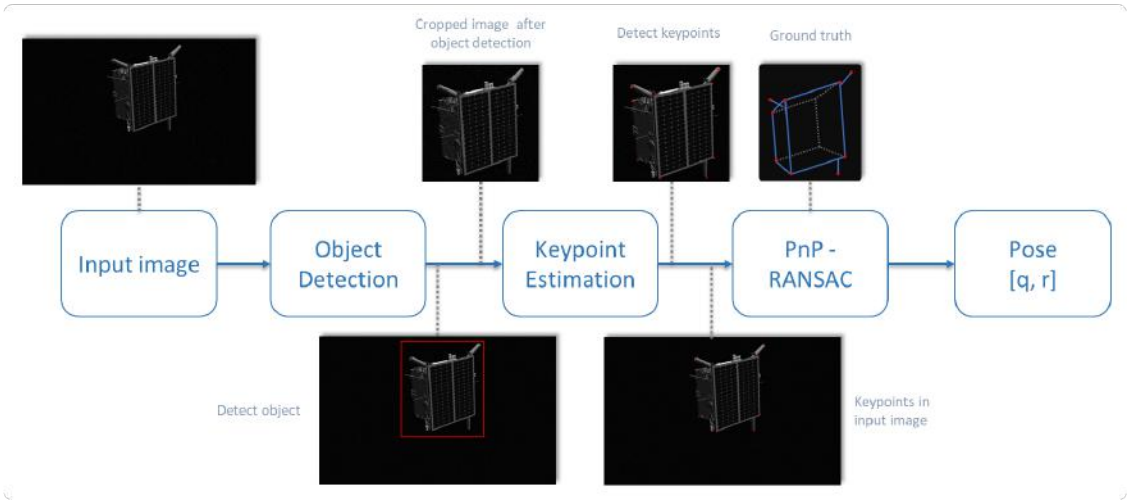


Figure 1: Keypoint-based Satellite Pose Estimation Framework

4.2 Keypoint regression

The bounding box in an image is cropped and provided as input to the keypoint regression framework. The inputs to the training the keypoint regression are the bounding box coordinates and the locations of the landmarks in the original image. Then crop the image and convert the landmark locations to the corresponding cropped image coordinates. Each landmark input has three columns, two for pixel coordinates *(x and y)* and one for the visibility. The landmark visibility value is set to either 0 or 1, depending on whether the landmark is visible in the image.

$$v_i = \begin{cases} 1 & \text{if visible to camera \& inside image frame} \\ 0 & \text{otherwise.} \end{cases}$$

We used HigherHRNet [11] to regress the 2D landmark locations and it uses a HRNet [12] as backbone. HigherHRNet provides an output at two different scales $1/4$ and $1/2$. We used the architecture that has 32-channels in the highest resolution feature maps. The output of the model depends on the number of landmarks in the particular model. For example, the SPEED model has 11 keypoints and the Soyuz model has 21 keypoints.

4.3 PnP + RANSAC

With the known 2D and 3D correspondences, PnP algorithm can be applied to compute the camera pose. There are possibilities of false correspondences among the derived keypoints, and to eliminate the outliers a RANSAC-based outlier rejection applied before estimating the camera pose. Unlike conventional sampling techniques that uses as much as data possible to obtain an initial solution, RANSAC generates solution by using the minimum number observations with the smallest set possible to estimate the underlying model parameters and then carry on to grow this set with consistent data inputs.

5 EXPERIMENTS

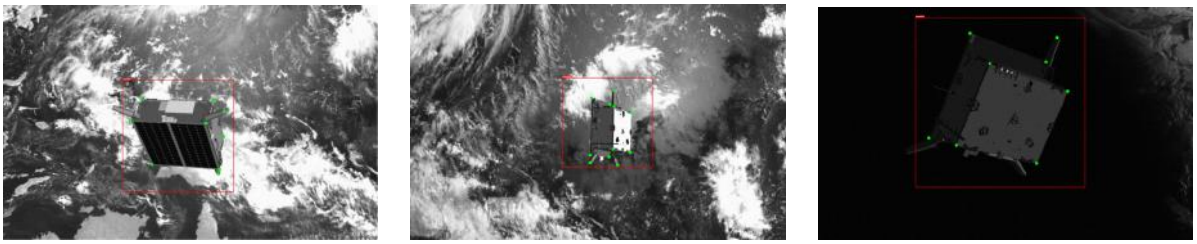
We conducted experiments on two different models from the SPEED dataset and the OrViS - Soyuz dataset. The models were trained using PyTorch on NVIDIA Quadro P4000 GPU. For training the model to detect bounding box, the input size of the image is set to 320 with a batch size of 8 and learning rate of 0.005, SGD with a momentum of 0.9, a weight decay regularization of 0.0005. During the training, the models are loaded with the pre-trained weights from COCO dataset and further tuned to identify the object using the bounding box coordinates. Input image augmentations added to make the model more robust. The training converges quickly within 20-30 epochs for both the models.

For training the model to regress the locations of the keypoints, the input image size is set to 640 with a batch size of 4 and an adam optimizer is used with 0.001 learning rate and 0.9 momentum. For keypoint regression, the training time is quite extensive and the image augmentation plays a key role in identifying the right keypoints. Image augmentation performed on the input image includes random rotation ($[-30°, 30°]$), random translation up to ($[-30\%, 30\%]$), coarse dropouts, Gaussian Noise, random brightness and contrast.

6 RESULTS

SPEED dataset

In the Pose Estimation Challenge, the geometry of the actual model kept as withheld information, different



(a) (b) (c)

Figure 2: Detected Bounding box and Keypoints using trained model for SPEED dataset

approaches were used across the competition entries either to find the true location or to overcome this. We used a structure-from-motion approach to estimate the actual locations of the interested keypoints. A Factor graph-based approach used to construct the relations between the keypoint locations in the images to the actual pose of the respective image [13]. We selected 10 images and manually identified the coordinates of the visible keypoints (out of 11 interested points) in each of these images.

Once the ground truth information is collected, the models are trained independently to detect the satellite and the respective keypoints. The trained models were used to inference the images in the SPEED test dataset and the estimated poses were uploaded to Kelvin's Pose Estimation Challenge post-competition submission portal. The results were summarized in table 1 and the current solution ranked next to the competition winner UniAdelaide, with the best score of 0.0096 and the real image score of 0.28973. Sample images from during inference, with the detected bounding boxes and the keypoints, are shown in fig. 2.

Rank	Name	Best Score	Real Image Score
1	UniAdelaide	0.0086	0.3634
*	STAR LAB keypoint method	0.0096	0.2897
2	EPFL_cvlab	0.0204	0.1040
3	pedro_fairspace (STAR LAB non-keypoint or direct method)	0.0554	0.1476
4	stanford_slab	0.0610	0.3221
5	Team_Platypus	0.0674	1.7117

Table 1: Top scores of Pose Estimation challenge

Soyuz dataset

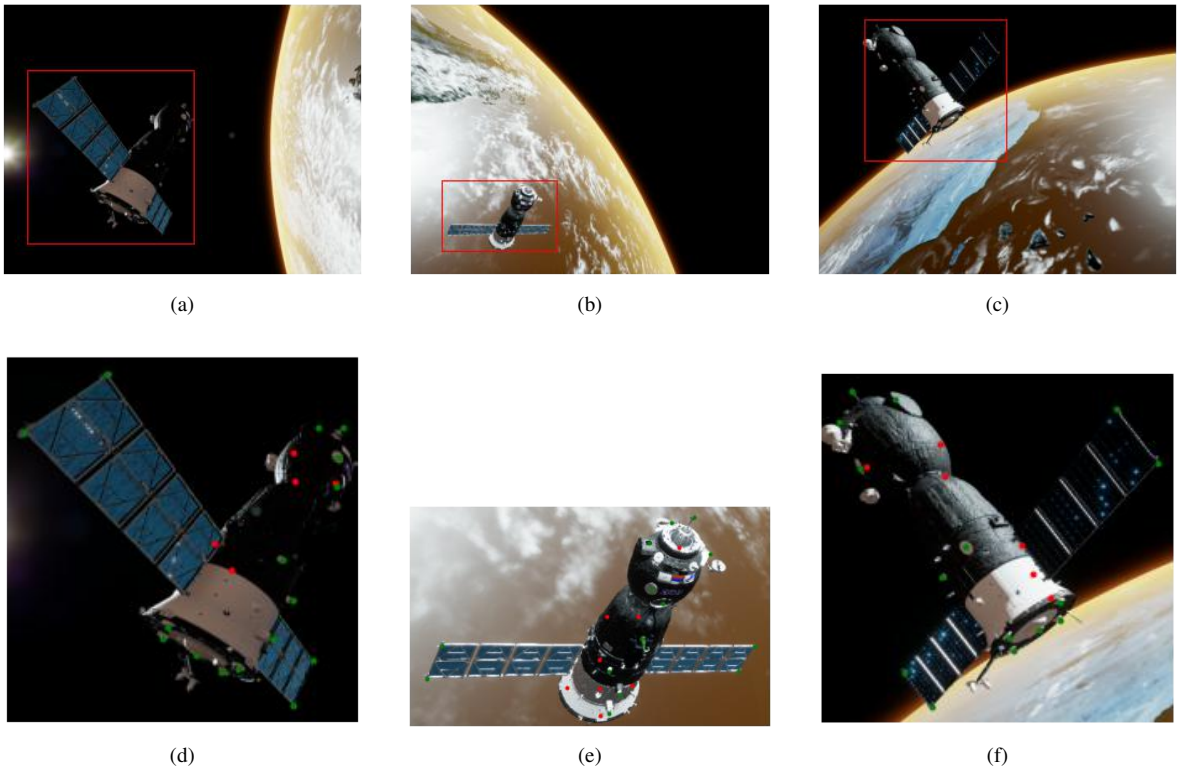
The Soyuz dataset from OrViS contain 5000 images, of which 10% reserved for testing and another 10% for validation. The results were recorded as the mean absolute location error, the mean angular error. For the keypoints-based approach the 21 keypoints were selected for the soyuz spacecraft.

The Tango spacecraft model in SPEED dataset is simple with 11 keypoints representing the target boundary limits and hence it is easier to compute the visibility of the keypoints by the simple intersection of planes. However, the Soyuz model has thousands of vertices and faces, and hence a simplified CAD model created as shown in fig. 3 to compute the visibility via ray tracing. Python package trimesh [14] used to perform ray tracing and to compute the visible points using the ground truth relative pose for each image. The ground truth samples of the bounding box locations and keypoint coordinates along with their visibility shown in fig. 4.

Figure 3: Simple Soyuz Model used for ray-tracing to compute points visibility (ground truth)

6.1 Comparison

Both the keypoint and the non-keypoints based approaches provide the advantages of its own. The keypoints-based framework is relatively simple and more scalable. Whereas, the direct (non-keypoint) framework provides robustness irrespective of the camera intrinsic parameters. The orientation resolution achievable with the direct approach depends on the number of bins used to encode the quaternion space during probabilistic soft classification. With the increase in the number of bins, the number of parameters in the trained model increases. For example, with



(a) (b) (c)

(d) (e) (f)

Figure 4: Samples of Bounding box and keypoint estimates for Soyuz dataset (green indicates visible keypoints $v_i = 1$ and red indicates occluded keypoints $v_i = 0$)

$64 \times 64 \times 64$ bins, a more accurate result for the SPEED dataset was achieved and it requires around 500M parameters [5]. However, the number of parameters for both object detection and the keypoint-based methods is around 50M-80M depending on the configurations. The results for the Soyuz dataset using both the direct and keypoint approach is summarized in table 2.

Method	Location Error	Angular Error
STAR LAB non-keypoint or direct method [5]	$0.8m$	$7.4°$
STAR LAB keypoint method	$0.3m$	$4.9°$

Table 2: Results for different approaches for Soyuz dataset

7 CONCLUSION

This work presented a keypoints-based deep-learning framework for spacecraft pose estimation employing different algorithms for both object detection and keypoint estimation while offering performance representing the state-of-the-art algorithm. The keypoints-based framework offers better accuracy with a minimum number of parameters than the direct approach. It provides the ability to easily modify the pipeline and update the components and test with state-of-the-art algorithms. The keypoint-based framework found to be more scalable than the direct approach.

7.1 Future work

Further experiments are inline to extensively test and validate the two approaches in the Ground-based hardware-in-the-loop experiments using a high-DOF testbed [15] at the STAR LAB, Surrey Space Centre. This testbed uses smaller laboratory robotic arm mounted on a traverser and implements the orbital dynamics into the robotic arm motion to simulate the close proximity motion of the servicer approaching the target. The testbed designed to operate on an open-source ROS framework is suitable for testings and validations of autonomous spacecraft GNC systems for small satellites. Additionally using state estimation filters during the navigation phase will marginalise the error in the estimates.

Acknowledgments

This work is supported by grant EP/R026092 (FAIRSPACE Hub) through UK Research and Inno-

vation (UKRI) under the Industry Strategic Challenge Fund (ISCF) for Robotics and AI Hubs in Extreme and Hazardous Environments.

References

[1] Llorente, J. S., Agenjo, A., Carrascosa, C., de Negueruela, C., Mestreau-Garreau, A., Cropp, A., & Santovincenzo, A. (2013). Proba-3: Precise formation flying demonstration mission. *Acta Astronautica*, *82*(1), 38–46.

[2] Bodin, P., Noteborn, R., Larsson, R., Karlsson, T., D'Amico, S., Ardaens, J. S., Delpech, M., & Berges, J.-C. (2012). The prisma formation flying demonstrator: Overview and conclusions from the nominal mission. *Advances in the Astronautical Sciences*, *144*, 441–460.

[3] IET, E. (2020). One satellite services another in orbit for the first time. https://eandt.theiet.org/content/articles/2020/04/one-satellite-has-serviced-another-in-orbit-for-the-first-time/

[4] Sharma, S., & D'Amico, S. (2019). Pose Estimation for Non-Cooperative Rendezvous Using Neural Networks, In *AIAA/AAS Space Flight Mechanics Meeting*, Ka'anapali, HI, IEEE. https://arxiv.org/abs/1906.09868

[5] Proenca, P. F., & Gao, Y. (2020). Deep Learning for Spacecraft Pose Estimation from Photorealistic Rendering, In *ICRA 2020: International Conference on Robotics and Automation*, (Virtual) Paris, France, IEEE. https://arxiv.org/abs/1907.04298

[6] Chen, B., Cao, J., Parra, A., & Chin, T. (2019). Satellite pose estimation with deep landmark regression and nonlinear pose refinement, In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. https://arxiv.org/abs/1908.11542

[7] Gerard, K. (2019). *Segmentation-driven satellite pose estimation* (Technical Report).

[8] Dhamani, N., Martin, G., Schubert, C., Singh, P., Hatten, N., & Akella, M. R. (2020). Applications of Machine Learning and Monocular Vision for Autonomous On-Orbit Proximity Operations, In *AIAA Scitech 2020 Forum*, Orlando, FL, American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2020-1376

[9] Harvard, A., Capuano, V., Shao, E. Y., & Chung, S.-J. (2020). Pose Estimation of Uncooperative Spacecraft from Monocular Images Using Neural Network Based Keypoints, In *AIAA Scitech 2020 Forum*, Orlando, FL, American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2020-1874

[10] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition, In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

[11] Cheng, B., Xiao, B., Wang, J., Shi, H., Huang, T. S., & Zhang, L. (2020). HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose Estimation, In *CVPR*.

[12] Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation, In *Cvpr*.

[13] Dellaert, F., & Beall, C. (2017). GTSAM:Ver-4.0. *Github*. https://github.com/borglab/gtsam

[14] Dawson-Haggerty et al. (2019). *Trimesh* (Version 3.2.0). https://trimsh.org/

[15] Hao, Z., Mavrakis, N., Proenca, P., Darnley, R. G., Fallah, S., Sweeting, M., & Gao, Y. (2019). Ground-based High-DOF AI and Robotics Demonstrator for In-Orbits Space Optical Telescope Assembly, In *Proceedings of the International Astronautical Congress*, Washington, US, International Astronautical Federation.