Close-range relative navigation with EPOS6D pose estimator*

Karl Martin Kajak[1][0000−0003−3029−0400] and Ksenia Klionovska[1][0000−0002−6739−3035]

DLR Oberpfaffenhofen, Münchener Str. 20, 82234 Weßling, Germany
{karl.kajak,ksenia.klionovska}@dlr.de

Abstract. The EPOS6D method [1] is used in a vision-based relative navigation setting to estimate the pose of an uncooperative target satellite in monocular camera images. EPOS6D is a pose estimation system composed of multiple sequentially acting components. The first stage is a neural network performing simultaneously object segmentation, part segmentation, and keypoint regression. The part segmentations and keypoints belong to specific fixed areas on the surface of the target object and allow establishing correspondences between 2D keypoints in the image and 3D keypoints on target object models, which is done via a Progressive-X implementation of PnP-RANSAC in the EPOS6D system. The EPOS6D method is applicable for target objects exhibiting symmetry or some other form of visual ambiguity that causes multiple pose hypotheses to be plausible. Previously, the EPOS6D method has been demonstrated on the T-LESS, YCB-V, and LM-O datasets from the general computer vision field. Using the method for pose estimation in space brings with it new challenges like large distance and lighting variations. Therefore, EPOS6D is tested on new datasets that reflect such spaceborne conditions, featuring synthetic and real images of a spacecraft exhibiting symmetry. The first focus of the experiments is on determining parameters and pose estimation system design aspects that allow training the network on synthetic images while actually developing it for use on the real camera images, or in other words, overcoming the "sim2real" problem. The second focus is on determining the accuracy and reliability potential of the system given the difficult visual conditions featured in the datasets. Finally, an attempt is made to try and initialize a visual tracker with the EPOS6D system in an attempt to improve the refresh rate of the pose estimations.

Keywords: Relative navigation · Pose estimation · Monocular camera · Convolutional neural networks · sim2real · Domain randomisation

1 Introduction

Active debris removal and in-orbit servicing activities of the future will require servicer spacecraft to know accurately the relative position and attitude of their

* This research was funded under EU H2020 MSCA ITN Stardust-R, grant agreement 813644.

client or target spacecraft. These targets are likely to not provide any assistance with this interaction and therefore lack visual markers, the ability to communicate navigation information, or the ability to stabilize their own orbit or attitude. Monocular cameras are an interesting sensor for vision-based relative navigation as they are ubiquitous, cheap, and high resolution.

Typically, a pose estimation system for relative navigation separates two tasks - initialization and tracking. Initialization refers to the task of estimating pose without prior estimates and therefore having a challengingly large search space. Tracking refers to the task of estimating pose with a prior estimate, therefore reducing the search space to a small neighborhood around the previous estimate. This usually means a lower computational cost and higher precision. In computer vision in general, convolutional neural networks (CNNs) have dominated the scene of visual perception since 2012, though now visual transformers are on the rise as well. CNNs have the potential to be flexible with respect to the specific appearance of the target spacecraft and the rest of the environmental conditions in camera images such as foreground and background variations, lighting condition variations, reflective materials, etc. Furthermore, if a generally suitable CNN-based system can be established, the development effort for each mission can be significantly streamlined due to not having to handcraft algorithms for each target object. This would be the case with techniques relying on classical image processing techniques that tend to focus on specific features like lines or corners.

One topic that has not yet settled when it comes to estimating pose with CNNs is what quantities specifically should be estimated with the network. Direct estimations have been attempted via discrete classification and regression of different rotation parameterisations such as Euler angles, quaternions, and axis-angle representations, for example. Most state-of-the-art solutions in general computer vision rely on regressing keypoints that belong to specific points on an object and then relating 2D keypoints in the image to the 3D keypoints on target object surfaces to obtain a pose estimate. One problem that affects the most solutions is ambiguous pose estimates due to some kind of visual characteristic in the image. This could be caused by the symmetry of the object as well as other factors such as adverse lighting conditions that hide unique features necessary to resolve a unique pose. EPOS6D is a CNN-based pose initialization system that has been designed with symmetric objects in mind while also having the potential to cope with other ambiguity-causing effects such as powerful lighting phenomena. This is due to the fact that the network is trained to recognize specific regions on an object in addition to recognizing the object on the whole. Utilizing the probabilistic correspondences of these regions and their corresponding keypoint estimates give rise to a capability to give at least plausible pose estimates in ambiguous conditions and unique pose solutions when the necessary discernible visual features are seen.

Due to the poor availability of real camera data it is desirable to train CNNs on synthetic data and have the resulting networks be applicable also to the real camera images. This issue is referred to as the "sim2real" problem in computer vision literature. A domain randomization approach has been integrated into the synthetic image generation procedure to force a CNN trained on this data to generalize to real camera images as well.

2 Contribution

The presented work documents experiments aimed at improving the domain generalization capability of the EPOS6D pose initializer. Preliminary results with default hyperparameters from the software repository of the authors of EPOS6D have not shown a great ability to generalize to the real camera images via training on domain randomized synthetic data.

Furthermore, the experiments explore the accuracy potential of the pose initialization given that the target object is at various distances from the camera. This is a crucial aspect that was not explored in the seminal work as it was only tested on T-LESS, YCB-V, and LM-O datasets, which do not feature distance variations as high as are visible in close proximity relative navigation conditions.

Lastly, the initializer is tested with a pose tracker to see how well it can initialize it and how they compare in terms of robustness towards visual conditions.

3 Methodology

The synthetic images of the spacecraft were rendered using the Cycles raytracer engine within the Blender software package. Variations of the spacecraft appearance were achieved with randomized texture, normal map, and material nodes. Background variations were also achieved with randomized texture nodes mixed with a representative orbital environment texture for an Earth orbiting satellite. Lighting variations are achieved via randomization of the direction and strength of the Sun light source as well as the camera exposure time. The real camera images were collected in the EPOS 2.0 facility at DLR Oberpfaffenhofen. They feature a representative approach maneuver from 17 meters up to 6 meters, finishing with a flyaround inspection maneuver at close range. To obtain preliminary results, the EPOS6D network was trained with a Resnet-50 backbone to extract features at image sizes 256x256. The features are forwarded to a DeepLabv3+ decoder for segmentation of the target spacecraft and its local regions as well as the centroid keypoints of the local regions. The training batch consists of 95400 synthetic images featuring domain randomization. The performance is the pose initializer is measured via pose accuracy histograms across the datasets. Pose accuracy is measured as the degrees error for the main symmetrical axis of the spacecraft (longitudinal axis) and the degrees error for the rotation about the symmetrical axis of the spacecraft (lateral axis). An illustrative image of the spacecraft is shown in Figure 1.

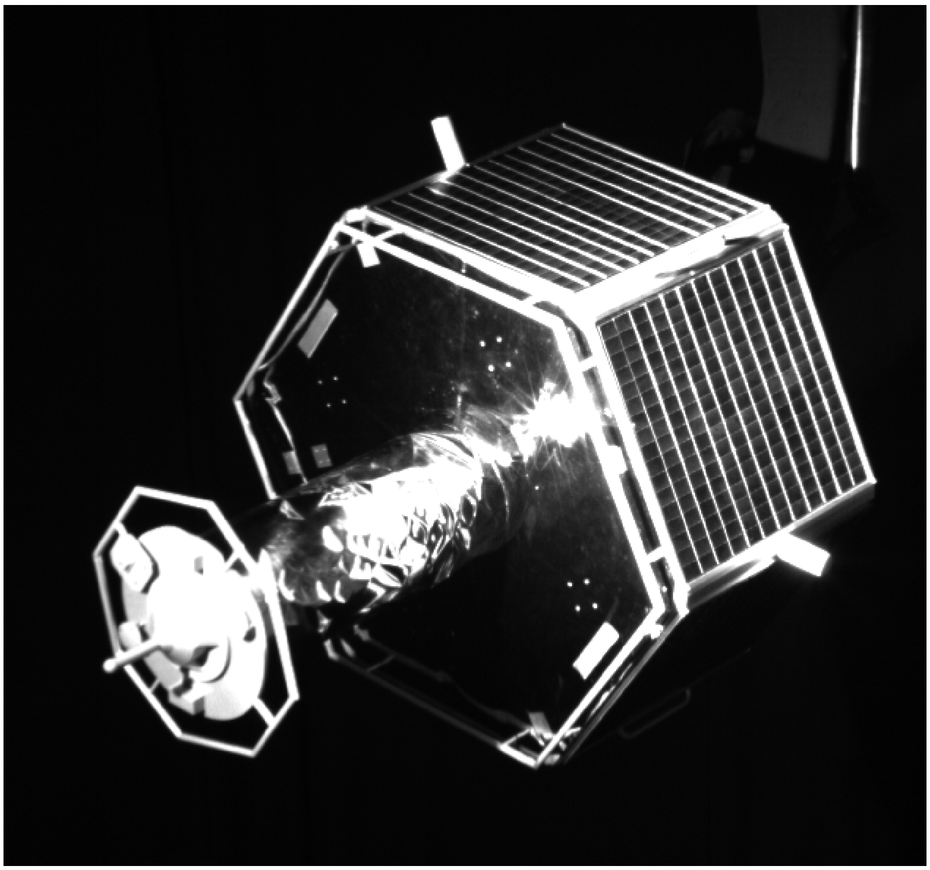

Fig. 1. An illustrative sample of how the spacecraft looks on the real camera images in EPOS 2.0 (cropped)

4 Preliminary results

Preliminary results are presented from evaluating the performance of the network on 300 synthetic images that were put aside from the training dataset to determine "maximum" performance as these images belong to the same domain as the training images. Figure 2 and 3 show the longitudinal and lateral error over this dataset, respectively.

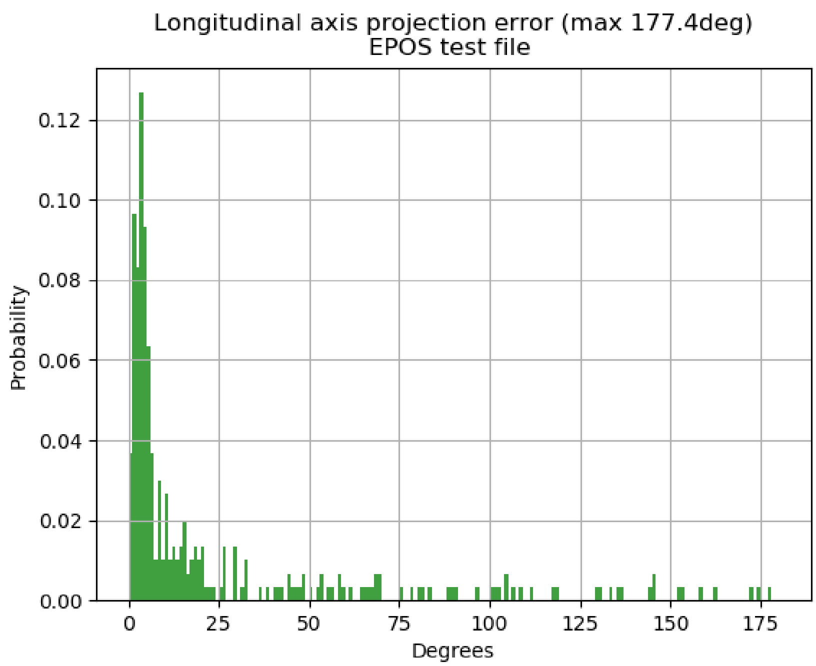

Fig. 2. A histogram of longitudinal error for the 300 image test dataset.

Preliminary results show that the initial parameters and design for the network are promising, but are lacking in a number of ways. Firstly, there is a non-negligible number of outlying pose estimates. Secondly, the single unique pose does not seem to be resolved more often than other plausible pose solutions.

5 Conclusions

Given the identified drawbacks of the initial parameters and network design, further experiments will focus on improving upon these and also, importantly, taking on the challenge of generalizing to real camera images from purely synthetic training data.

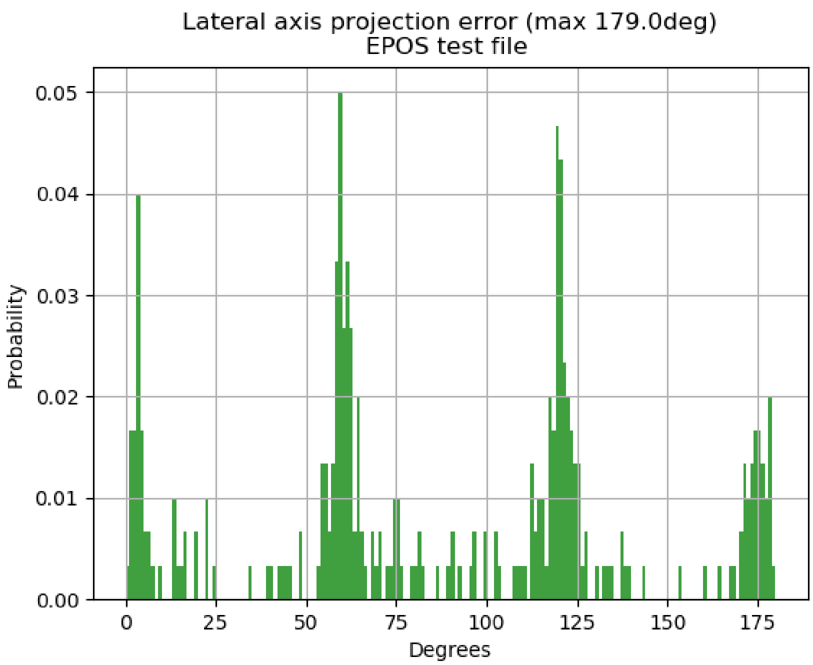

Fig. 3. A histogram of lateral error for the 300 image test dataset.

References

1. Hodan, T., Baráth, D., Matas, J.: EPOS: Estimating 6D pose of objects with symmetries. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 11700–11709 (2020)