

# ROBOTIC TESTBED FOR RENDEZVOUS AND OPTICAL NAVIGATION: MULTI-SOURCE CALIBRATION AND MACHINE LEARNING USE CASES

Tae Ha Park\*, Juergen Bosse†, Simone D'Amico‡

This work presents the most recent advances of the Robotic Testbed for Rendezvous and Optical Navigation (TRON) at Stanford University - the first robotic testbed capable of validating machine learning algorithms for spaceborne optical navigation. The TRON facility consists of two 6 degrees-of-freedom KUKA robot arms and a set of Vicon motion track cameras to reconfigure an arbitrary relative pose between a camera and a target mockup model. The facility includes multiple Earth albedo light boxes and a sun lamp to recreate the high-fidelity spaceborne illumination conditions. After the overview of the facility, this work details the multi-source calibration procedure which enables the estimation of the relative pose between the object and the camera with millimeter-level position and millidegree-level orientation accuracies. Finally, a comparative analysis of the synthetic and TRON simulated imageries is performed using a Convolutional Neural Network (CNN) pre-trained on the synthetic images. The result shows a considerable gap in the CNN's performance, suggesting the TRON simulated images can be used to validate the robustness of any machine learning algorithms trained on more easily accessible synthetic imagery from computer graphics.

## INTRODUCTION

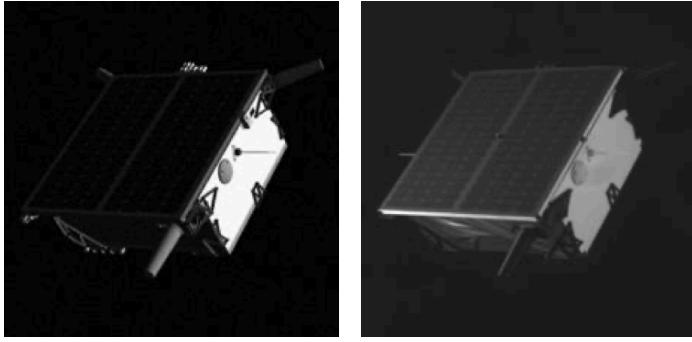
The vision-only navigation of a spacecraft about noncooperative Resident Space Objects (RSO) is an enabling technology for future on-orbit servicing and debris removal missions. Unlike those based on complex sensors such as Light Detection and Ranging (LIDAR) or stereovision, monocular navigation systems utilize a commercially available low Size, Weight, Power, and Cost (SWaP-C) camera, making it an attractive choice of sensor due to its low mass and power requirements. The key component of monocular navigation is to determine the pose (i.e., position and orientation) of the target relative to the servicer's camera based on a single or a sequence of images. The conventional approach is to first extract and process salient features such as points,<sup>1</sup> edges,<sup>2,3</sup> scale-invariant features such as SIFT,<sup>4</sup> SURF<sup>5</sup> and ORB<sup>6</sup> features of a spacecraft, or landmark features such as craters of an asteroid.<sup>7</sup> These features are then compared with those of the available target 3D model to compute the 6D pose.<sup>7–12</sup> Recently, Machine Learning (ML) techniques based on Convolutional Neural Networks (CNN) have been developed to replace the feature extraction step with superior performance.<sup>13–17</sup> However, unlike image processing-based methods, a CNN must be trained on a large set of target images with accurate pose labels. While the data-hungry nature of the training is a ubiquitous challenge for any ML applications, it is especially difficult and outright impractical in spaceborne applications to acquire a large quantity of images of interested targets in

---

\*Ph. D. Candidate, Department of Aeronautics & Astronautics, Stanford University, Stanford, CA 94305.

†Managing Director, Robo-Technology GmbH, Benzstrasse 12, D-82178 Puchheim, Germany.

‡Associate Professor, Department of Aeronautics & Astronautics, Stanford University, Stanford, CA 94305.

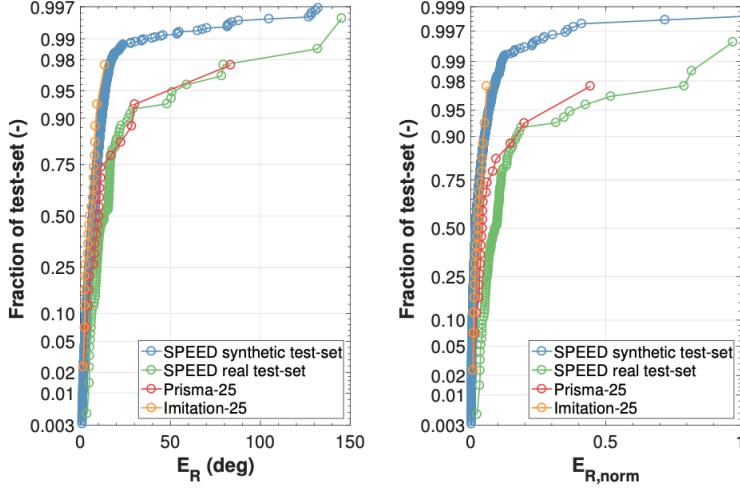


**Figure 1.** SPEED synthetic (*left*) and simulated (*right*) cropped images with the pose label estimated from the TRON testbed.

various space environments with accurate pose labels. Therefore, existing works rely on synthetic images generated from a computer graphics renderer such as OpenGL<sup>13–15, 17</sup> or Blender<sup>11</sup> to train and test the CNN subsystems.

Unfortunately, the synthetic renderers cannot faithfully replicate various illumination and noise artifacts present in the spaceborne imagery. Naturally, CNNs trained with synthetic images alone would overfit to the features inherent to the synthetic imagery and thus have degraded performance on the spaceborne imagery.<sup>14, 15, 17, 18</sup> Therefore, in order for a CNN-based system to be deployed to space missions: 1) it must be trained to be robust against various adversarial conditions in space such as high contrast, extreme shadowing and low signal-to-noise ratio, and 2) its robustness must be validated on ground with no access to target spaceborne images. Previous works have attempted to address the robust training with extensive data augmentation during training,<sup>14, 16</sup> but relatively fewer works or efforts have been dedicated to the issue of on-ground validation of CNNs.<sup>13</sup> The most promising method involves a robotic testbed that is capable of re-creating various space environmental conditions and configuring the camera and the target model to achieve the desired relative pose with high accuracy. Such facility would allow one to obtain the quasi-spaceborne imagery of an arbitrary quantity and characteristics with statistical distribution completely different from synthetic images. Such an imagery can then be used to evaluate the robustness of the CNN trained with synthetic images or any other vision-based navigation algorithms developed based on them.

One example of such approach is the Spacecraft Pose Estimation Dataset (SPEED),<sup>19</sup> which was made publicly available in 2019. SPEED contains 15,000 synthetic images of the Tango spacecraft from the PRISMA mission<sup>8, 20</sup> and 300 simulated images of the full-scale Tango mockup model captured from the Robotic Testbed for Rendezvous and Optical Navigation (TRON) at the Stanford’s Space Rendezvous Laboratory (SLAB). As seen in Figure 1, a simulated image describes identical geometric features of the target yet has fundamentally different visual characteristics compared to its synthetic counterpart. In fact, the result of the Satellite Pose Estimation Competition (SPEC), co-hosted by SLAB and the Advanced Concepts Team (ACT) of the European Space Agency (ESA) in 2019, shows virtually all the top-performing CNNs trained on synthetic images have degraded performances on simulated images.<sup>15</sup> The similar trend is also described in Figure 2, which shows that the Spacecraft Pose Network (SPN) model trained on SPEED synthetic images have worse attitude predictions on SPEED simulated (real) test set and PRISMA-25, which consists of 25 spaceborne images from the rendezvous phase of the PRISMA mission. Interestingly, the degraded yet comparable performances on the simulated and spaceborne images suggest that the simulated images from a robotic testbed with the capabilities of TRON can thus be used to evaluate the robustness



**Figure 2.** (Left) Cumulative distribution of rotation error of SPN<sup>13</sup> when tested on SPEED synthetic test set, real test set, PRISMA-25, and Imitation-25 datasets. (Right) Same distribution of the rotation error normalized by inter-spacecraft separation and camera field-of-view. Figures taken from Ref [18].

of a CNN for spaceborne applications. However, the simulated imagery of SPEED is currently restricted to 300 images with extremely restricted pose distribution and variety of illumination conditions. Therefore, a significant upgrade must be made to validate the performance on a wide range of navigation scenarios.

Several other laboratories have constructed similar testbeds to simulate vision-only closed-loop navigation and control algorithms. Some examples include ASTROS at the Georgia Institute of Technology,<sup>21,22</sup> POSEIDYN at the Naval Postgraduate School,<sup>23</sup> and M-STAR at California Institute of Technology.<sup>24</sup> These facilities commonly employ air-bearing platforms on a flat epoxy or granite floor with thrusters and actuators to simulate the spacecraft movement and Vicon motion-tracking cameras to provide ground-truth pose labels. While these testbeds excel at simulating actual spacecraft movement given maneuver commands, none of them are tailored for ML applications, as the capability of efficiently reconfiguring the arbitrary pose commands at large quantity has never been showcased. Recently, the GNC Rendezvous, Approach and Landing Simulator (GRALS) testbed at the European Space Research and Technology Centre (ESTEC), a facility comprising a ceiling-mounted KUKA robotic arm and Vicon motion track cameras, was used to generate 100 simulated images of 1:25 mockup of the Envisat satellite.<sup>25</sup> However, similar to the earlier generation of TRON, the target is mounted on a static tripod, severely restricting the image acquisition from different viewpoints.

The first contribution of this paper is the introduction of the next generation TRON facility at SLAB, the first testbed capable of accurately reconfiguring an arbitrary relative pose with high-fidelity space-like illumination conditions. Unlike other facilities, TRON includes two KUKA 6 degrees-of-freedom (DOF) robot arms<sup>26</sup> respectively holding a camera and a lightweight, reduced-scale model of the target RSO (see Figure 3). One robot is installed on a ceiling-mounted linear rail running through the facility; therefore, compared to its previous generation with only one robot arm,<sup>13,27</sup> the facility as a whole provides total 13 DOF and allows to take images of the target from the full orientation space and the distance between two objects up to 6 meters. To the authors'

knowledge, this capability of TRON is only rivaled by the European Proximity Operations Simulator (EPOS) at DLR, which also consists of two 6DOF KUKA robot arms to simulate the rendezvous and proximity operations in space.<sup>28</sup> In addition, the TRON facility is equipped with 10 Earth albedo light boxes around the room simulating the diffuse light<sup>27</sup> and a metal halide arc lamp simulating the direct sunlight.

The second contribution of this paper is the multi-source calibration procedure of TRON using two independent measurement systems: 1) KUKA telemetry which provides the end-effector poses, and Vicon motion track system which provides the objects' poses via tracking the infrared (IR) markers attached to them. Given rigid fixtures of both the camera and the target onto the respective end-effectors, the calibration amounts to solving for each measurement source the Robot/World Hand/Eye (RWHE) calibration problem<sup>29</sup> associated with either measurement source. This allows one to reconstruct the target's pose based on either KUKA or Vicon measurements. The final pose estimate is obtained by fusing the reconstructed poses in a Bayesian framework, which helps reduce the effect of any bias or noise present in either measurements. Moreover, a criterion is developed to reject any bad measurements reported by Vicon. The result shows that the calibrated poses of the target relative to the camera achieve on average sub-degree orientation and millimeter-level position accuracy at close range.

The third contribution of this paper is the comparative analysis of simulated images under a variety of illumination settings that the TRON facility is capable of recreating using its albedo boxes and the sun lamp. Specifically, a CNN pre-trained on SPEED synthetic training images is used on pairs of synthetic and simulated images with shared pose labels and aligned directions of the light source. A considerable performance gap is observed between two imageries, with far worse performance on the model illuminated with the sun lamp and viewed from certain directions. This suggests that the TRON simulated images exhibit a significant domain gap against the synthetic training images and thus can be used as a good database for validating a CNN's robustness across different domains.

This paper is organized as follows. It first provides an overall description of various components of the TRON testbed. Then, it describes the full calibration procedure and the data fusion mechanism to enable accurate pose label generation. An experiment calibration is run to show the accuracy reported by the testbed, and it ends with a comparative analysis of the synthetic and simulated image qualities using a pre-trained CNN.

## NOTATIONS

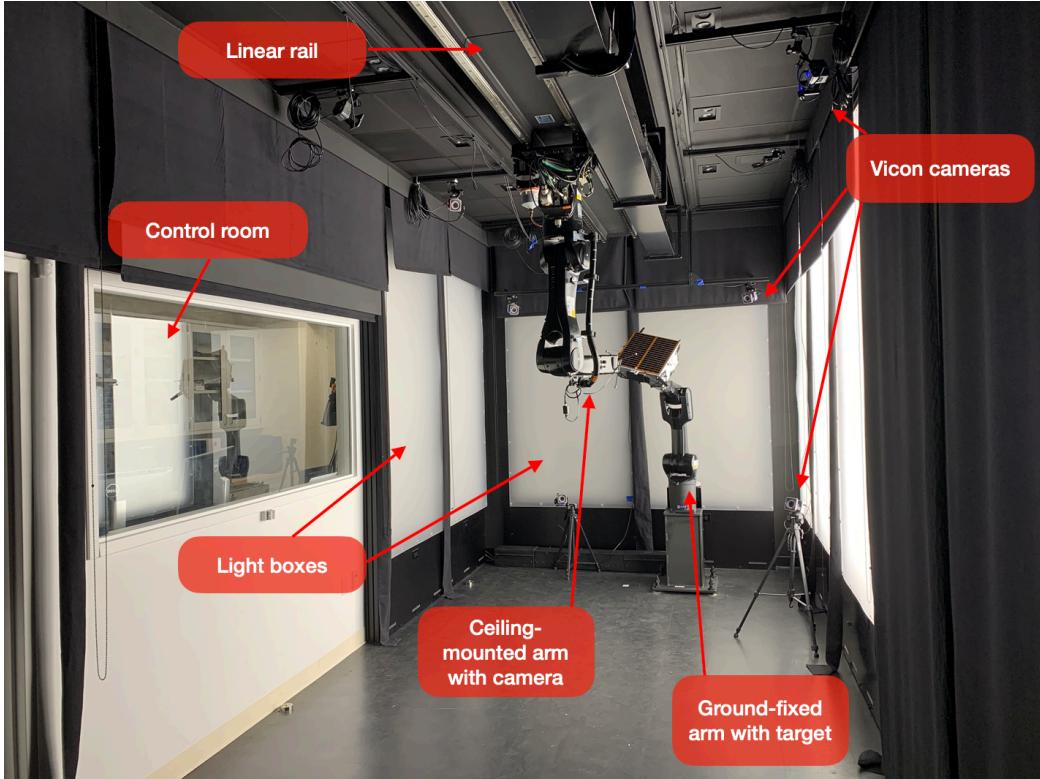
In this work,  $\mathbf{x}_A \in \mathbb{R}^3$  denotes a 3D vector expressed in a reference frame  $A$ , and  $\mathbf{x}_A^h = [\mathbf{x}_A^\top \ 1]^\top$  is a homogeneous vector extension of  $\mathbf{x}_A$ . Given two reference frames  $A$  and  $B$ , a point  $\mathbf{x}_B$  can be equivalently expressed in  $A$  via the following rigid transformation,

$$\mathbf{x}_A = \mathbf{R}_{BA}\mathbf{x}_B + {}^A\mathbf{t}_{AB}, \quad (1)$$

where  $\mathbf{R}_{BA} \in \mathbb{R}^{3 \times 3}$  is an orthonormal rotation matrix aligning  $B$  to  $A$ , and  ${}^A\mathbf{t}_{AB}$  is a translation vector from the origin of  $A$  to that of  $B$  expressed in  $A$ . Equation 1 is equivalent to the following transformation of homogeneous vectors,

$$\mathbf{x}_A^h = \begin{bmatrix} \mathbf{R}_{BA} & {}^A\mathbf{t}_{AB} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \mathbf{x}_B^h = \mathbf{T}_{BA}\mathbf{x}_B^h, \quad (2)$$

where  $\mathbf{T}_{BA} \in \mathbb{R}^{4 \times 4}$  is a combined roto-translation or transformation matrix between  $B$  and  $A$ .



**Figure 3.** TRON simulation room and its components.

## TRON FACILITY DESCRIPTION

The TRON facility at the Space Rendezvous Laboratory (SLAB) of Stanford University, visualized in Figure 3, includes a control room and an  $8 \times 3 \times 3$  [m] simulation room which consists of various components and machineries to 1) simulate the vision-based rendezvous trajectory of a servicer spacecraft with a camera to its target RSO, and 2) emulate the high-fidelity spaceborne illumination conditions to maximize the realism in the images captured by the camera. This section provides a high-level overview of the components of the facility enabling the above two goals. Note that the facility has seen an extensive upgrade to both its hardware and software capabilities since its original descriptions in Refs [13], [18].

### Pose Reconfiguration & Annotation

Given a relative pose to be achieved between the servicer's camera and the RSO model, including satellites, debris, and even celestial bodies such as asteroids and the landing sites, the pose reconfiguration is achieved by simultaneously controlling two 6 DOF KUKA robotic arms,<sup>26</sup> respectively holding a camera and a lightweight, reduced-scale mockup model of the RSO at their end-effectors (see Figure 3). The robot holding a camera is installed onto a ceiling-mounted linear axis rail, providing an additional DOF along the facility and up to approximately 6 meters of separation between the objects along the linear rail. The facility as a whole thus provides total 13 DOF and allows to take images of the target model with the orientation distribution covering the full  $SO(3)$  space. To the authors' knowledge, this capability is currently unavailable in any other similar testbeds except



**Figure 4. (Left) Two albedo light boxes activated. (Right) Sun lamp activated.**

in the EPOS facility at DLR. The desired RSO’s model can be manufactured with two mounting spots at the opposite sides, so that at each mounting configuration, only half of the orientation space is viewed by the camera in order to prevent the robot arm from ever blocking its sight.

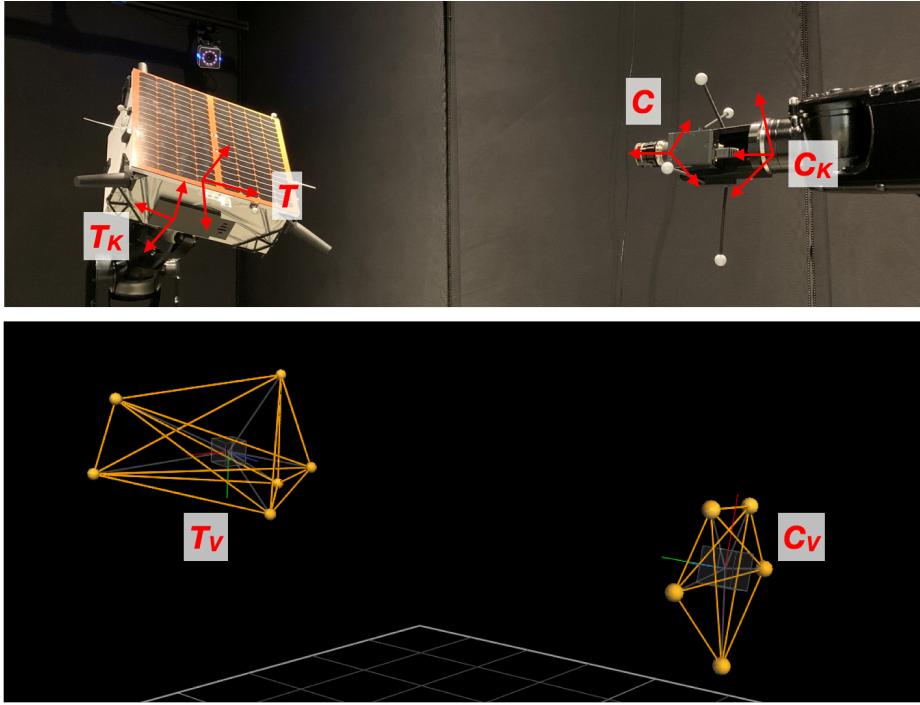
In order to track the movements of both objects, the facility includes 12 Vicon Vero cameras that track the IR markers attached to the objects.<sup>30</sup> The Vicon tracker software attaches a reference frame to a set of IR markers associated with each object and is capable of reporting its real-time position and orientation within the facility. Independent from the external measurements provided by Vicon, the KUKA system also provides the telemetry of the poses of both arms’ end-effectors in real-time based on their internal joint angles. These two sources of measurements are later jointly used to calibrate the facility, so that a user can retrieve the most accurate estimate of the pose between the camera and the target mockup model in any arbitrary configuration of the robot arms.

### High-Fidelity Illumination Condition

In order to physically recreate high-fidelity space illumination conditions, the TRON facility is equipped with 10 light boxes around the walls to simulate the diffuse light of Earth albedo.<sup>27</sup> A light box consists of a diffuser plate covering hundreds of Light Emitting Diodes (LEDs) arranged in strips that can be regulated in color and intensity. The light boxes are rigorously calibrated to output maximally uniform radiance across the diffuser plates consistent with Earth albedo in Low Earth Orbit (LEO).<sup>31</sup> The facility also includes a metal halide arc lamp capable of simulating a direct sunlight. Figure 4 illustrates the operation of both devices and a stark contrast of the effects they cast onto a model. All ambient light sources, including the deactivated light boxes and the windows, are covered with light-absorbing black commando curtains during operations to maximize the effect of diffuse and direct light.

### SINGLE-SOURCE CALIBRATION

The calibration of TRON aims to enable the estimation of the pose between the camera and the target RSO given measurements from KUKA and/or Vicon. This section first formulates the calibration problem based on a set of measurements from a single source: KUKA or Vicon. Then, it provides the descriptions of the full calibration procedure.



**Figure 5.** (Top) End-effector ( $C_K, E_K$ ) and object ( $C, T$ ) reference frames. (Bottom) Object reference frames assigned by Vicon ( $C_V, T_V$ ), visualized in the Vicon tracker software as RGB triads.

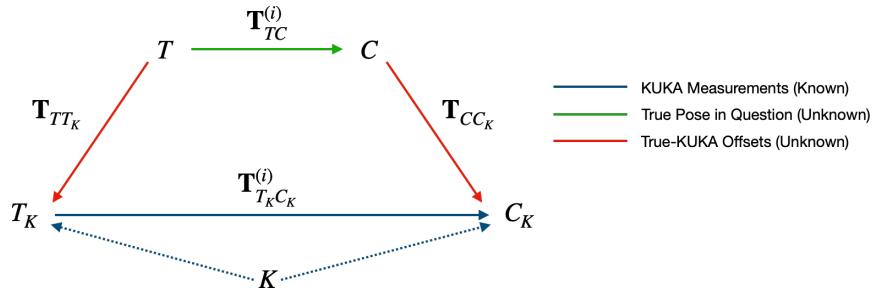
## Reference Frames

The following reference frames used in a single-source calibration are defined below and illustrated in Figure 5.

- *Camera true reference frame* ( $C$ ), whose  $z$ -axis is along the camera boresight and its  $xy$ -axes form the image plane.
- *Target true reference frame* ( $T$ ), defined according to the 3D model of the target.
- *KUKA end-effector frames* ( $C_K, T_K$ ), defined respectively for the end-effectors of the camera- and the target-holding robot arms.
- *Global KUKA reference frame* ( $K$ ), fixed at an arbitrary location on the ground.
- *Vicon object frames* ( $C_V, T_V$ ), defined respectively by the Vicon tracker software based on the set of IR markers associated with each object.
- *Global Vicon reference frame* ( $V$ ), also defined by the Vicon tracker software to be fixed at an arbitrary location within the facility.

## Single-Source Calibration Problem

Figure 6 visualizes the calibration problem in a KUKA-only setting. The goal is to estimate the true pose between the camera and the target,  $T_{TC}$ . Note that KUKA provides the measurements



**Figure 6. Diagram of the KUKA-only calibration problem.**

$(\mathbf{T}_{T_KK}, \mathbf{T}_{C_KK})$ , which denote the rigid transformations between the global KUKA frame and both robots' respective end-effector frames. Equivalently, the measurement provides the relative pose between two end-effectors,  $\mathbf{T}_{T_KC_K}^{(i)} = (\mathbf{T}_{C_KK}^{(i)})^{-1}\mathbf{T}_{T_KK}^{(i)}$ ,  $i = 1, \dots, N$ , for  $N$  calibration data points. There is also an offset between an end-effector and the object it holds, denoted by  $(\mathbf{T}_{T_KT}, \mathbf{T}_{C_KC})$ , that are constant regardless of robot configurations due to rigid fixture of an object onto its end-effector. Then, it is obvious from Figure 6 that if the user has a knowledge of  $(\mathbf{T}_{T_KT}, \mathbf{T}_{C_KC})$ , the true pose in question can be recovered via:

$$\tilde{\mathbf{T}}_{TC,K}^{(i)} = \mathbf{T}_{C_KC}\mathbf{T}_{T_KC_K}^{(i)}\mathbf{T}_{T_KT}^{-1}, \quad (3)$$

where  $\tilde{\mathbf{T}}_{TC,K}$  indicates that the pose has been estimated based on the KUKA measurements. Note that the setting is completely identical in case of Vicon-only calibration, in which given the measurements  $\mathbf{T}_{T_VC_V}^{(i)}, i = 1, \dots, N$ , knowing the constant offsets  $(\mathbf{T}_{T_VT}, \mathbf{T}_{C_VC})$  allows one to recover the true pose via

$$\tilde{\mathbf{T}}_{TC,V}^{(i)} = \mathbf{T}_{C_VC}\mathbf{T}_{T_VC_V}^{(i)}\mathbf{T}_{T_VT}^{-1}, \quad (4)$$

where  $\tilde{\mathbf{T}}_{TC,V}$  indicates that the pose has been estimated based on the Vicon measurements. Henceforth, the subscript denoting the measurement source ( $K, V$ ) is omitted for notational simplicity unless noted otherwise.

The single-source calibration problem now amounts to solving for the constant offsets  $(\mathbf{T}_{T_KT}, \mathbf{T}_{C_KC})$  or  $(\mathbf{T}_{T_VT}, \mathbf{T}_{C_VC})$ . Assuming one can recover the true poses for a limited amount of samples in a controlled calibration setting, denoted  $\mathbf{T}_{TC,cal}^{(i)}, i = 1, \dots, N$ , then Equations 3 and 4 can be rearranged to

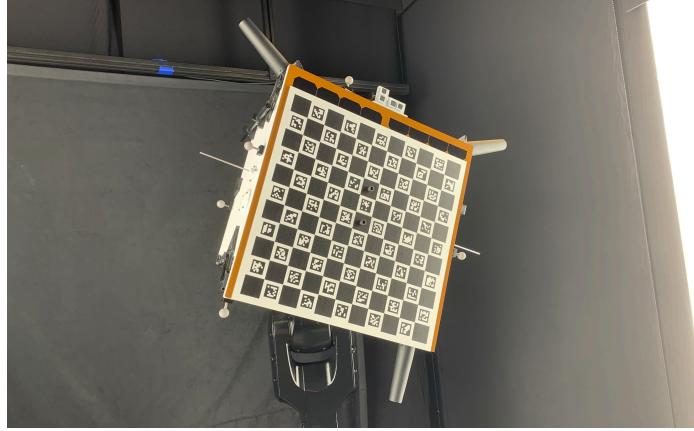
$$\mathbf{T}_{TC,cal}^{(i)}\mathbf{T}_{T_S T} = \mathbf{T}_{C_S C}\mathbf{T}_{T_S C_S}^{(i)}, i = 1, \dots, N, \quad (5)$$

where  $S \in \{K, V\}$  denotes the measurement source, and  $\mathbf{T}_{T_S T}, \mathbf{T}_{C_S C}$  are the only unknowns. Equation 5 is known as the Robot/World Hand/Eye (RWHE) calibration problem whose solution is well studied in literature.<sup>29,32,33</sup>

In summary, the calibration procedure to solve the single-source RWHE problem in Equation 5 consists of two steps: solve for  $\mathbf{T}_{TC,cal}$ , then solve the RWHE problem in Equation 5 for each given measurement source. Next sections describe how these two steps are resolved.

### Solving $\mathbf{T}_{TC,cal}$

First, the true pose can be accurately estimated using a known calibration camera and a pattern such as chessboard, asymmetric circle grids, or ChArUco which combines unique ArUco markers<sup>34</sup>



**Figure 7.** Calibration setup with a ChArUco pattern board on the model.

and a chessboard. In this work, a ChArUco pattern board is used to solve  $\mathbf{T}_{TC,cal}$ . Compared to other options, a ChArUco pattern offers especially flexible choices in terms of its placement with respect to the camera, as the presence of unique ArUco markers allows an easy identification of a partial set of the chessboard corners visible from the images.

As shown in Figure 7, a ChArUco pattern board of known dimensions is manufactured and fixed at the known location over a flat surface of the model such as the solar panel, so that the 3D coordinates of the pattern features (i.e., chessboard corners) are known in  $T$  reference frame. Then, one can simply solve the Perspective-n-Point (PnP)<sup>35</sup> problem to compute  $\mathbf{T}_{TC,cal}^{(i)}$  for  $i = 1, \dots, N$ , which involves minimizing the following reprojection error:

$$\min_{\mathbf{T}_{TC}^{(i)}, i=1, \dots, N} \frac{1}{M} \sum_{i=1}^N \sum_{j=1}^{M_i} \|\Pi(\mathbf{T}_{TC}^{(i)}, \mathbf{X}_j, \mathbf{K}, \mathbf{d}) - \mathbf{x}_j\|^2, \quad (6)$$

where  $\Pi(\cdot)$  is a projection operator,  $\mathbf{X}_j$  is the 3D location of the  $j^{\text{th}}$  feature,  $\mathbf{x}_j$  is the 2D location of the detected  $j^{\text{th}}$  feature in the image plane,  $\mathbf{K}$  contains the camera's intrinsic parameters,  $\mathbf{d}$  contains the lens distortion parameters, and  $M = \sum_{i=1}^N M_i$  is the total number of visible features in all images. It is also possible to simultaneously perform camera calibration by jointly minimizing  $(\mathbf{T}_{TC}, \mathbf{K}, \mathbf{d})$  in Equation 6. In this case, the usage of a ChArUco pattern is also advantageous, as the flexibility in its positioning facilitates the detection of the chessboard corners along the edges and corners of the image plane, which increases the quality of the estimated lens distortion parameters.

The detection of the ArUco markers and the chessboard corners is performed using the `aruco` library of OpenCV 4.5.2<sup>†</sup>.

### Solving RWHE Problem

Once  $\mathbf{T}_{TC,cal}$  are available from solving Equation 6, the RWHE problem in Equation 5 for both Vicon and KUKA measurements can be solved using any existing algorithms based on iterative nonlinear least-squares optimization,<sup>29</sup> Kronecker product of matrices,<sup>32</sup> dual quaternions,<sup>33</sup> and so

---

<sup>†</sup>[https://github.com/opencv/opencv\\_contrib](https://github.com/opencv/opencv_contrib)

on. This work frames the RWHE as the following nonlinear least-squares problem,<sup>29</sup>

$$\min_{\mathbf{T}_{T_S T}, \mathbf{T}_{C_S C}} \sum_{i=1}^N \|\mathbf{T}_{T_C, \text{cal}}^{(i)} - \mathbf{T}_{C_S C} \mathbf{T}_{T_S C_S}^{(i)} \mathbf{T}_{T_S T}^{-1}\|_F^2, \quad S \in \{K, V\}, \quad (7)$$

where  $\|\cdot\|_F$  denotes a Frobenius norm. In general, the rotation matrices  $(\mathbf{R}_{T_S T}, \mathbf{R}_{C_S C})$  can be parametrized using a quaternion, Euler angles, or the matrices themselves can be used directly in the optimization with proper constraints. This work uses the vector representation  $\mathbf{r}$  of a rotation matrix  $\mathbf{R}$  related by the Rodrigues formula, i.e.,

$$\mathbf{R}(\mathbf{r}) = \mathbf{I}_{3 \times 3} \cos \theta + (1 - \cos \theta) \mathbf{u} \mathbf{u}^\top + [\mathbf{u}]_\times \sin \theta, \quad (8)$$

where  $\theta = \|\mathbf{r}\|$ ,  $\mathbf{u} = \mathbf{r}/\theta$ , and  $[\mathbf{u}]_\times$  is a skew-symmetric matrix of  $\mathbf{u}$ . This parametrization requires no constraint and is also the adopted parametrization in OpenCV, which facilitates the interface with the OpenCV functionalities.

In summary, Equation 7 can be expanded to

$$\min_{\mathbf{r}_{TT_S}, \mathbf{t}_{T_S T}, \mathbf{r}_{C_S C}, \mathbf{t}_{CC_S}} \sum_{i=1}^N \|\mathbf{T}_{T_C, \text{cal}}^{(i)} - \begin{bmatrix} \mathbf{R}(\mathbf{r}_{C_S C}) & \mathbf{t}_{CC_S} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \mathbf{T}_{T_S C_S}^{(i)} \begin{bmatrix} \mathbf{R}(\mathbf{r}_{TT_S}) & \mathbf{t}_{T_S T} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix}\|_F^2, \quad (9)$$

which is solved using the Levenberg-Marquardt algorithm available in MATLAB's `lsqnonlin` function<sup>†</sup>.

## DATA FUSION

Once the offsets  $(\mathbf{T}_{T_K T}, \mathbf{T}_{C_K C}, \mathbf{T}_{T_V T}, \mathbf{T}_{C_V C})$  are estimated from the KUKA-based and Vicon-based RWHE calibration problems, one can use either Equation 3 or 4 to estimate  $\tilde{\mathbf{T}}_{TC}$  for any new data samples. In reality, one can also fuse the results from both measurement sources to help mitigate any bias or noise introduced to the offsets estimated in a single-source measurement setting. In contrast to the single-source calibration, a data fusion approach can take into account the fact that the reference IR markers cannot always be reliably tracked by multiple Vicon cameras, as a number of them inevitably fall into blind spots created by the robot arms and the target model as they move. The condition is exacerbated by the noise introduced by the reflective surfaces of the RSO model and the robot arm components. On the other hand, the KUKA measurements are reported with consistent accuracy based on the internal joint angles of the robot arms regardless of the end-effector pose in the room. Therefore, the limitation of Vicon necessitates a criterion to reject any outlying Vicon measurements, in which case the pose estimated from the KUKA measurement,  $\tilde{\mathbf{T}}_{TC, K}$ , is used alone as the final pose label. The following sections describe the data fusion strategy and the rejection criterion.

### Bayesian Data Fusion

This work employs a Bayesian approach to data fusion<sup>36</sup> which utilizes Bayes' theorem and Maximum A Posteriori (MAP) estimation of the posterior state. The Bayesian framework allows one to estimate the true state  $X$  as more observations  $Z$  become available. Let the probability density function  $p(Z | X)$  denote the likelihood function encoding the probabilistic information

---

<sup>†</sup><https://www.mathworks.com/help/optim/ug/lsqnonlin.html>

contained in the random variable  $Z$  about  $X$ . By Bayes' theorem, the *a posteriori* probability that  $X = x$  given the measurement  $Z = z$  can be expressed as

$$p(X = x | Z = z) \propto p(Z = z | X = x)p(X = x) \quad (10)$$

where  $p(X = x)$  denotes the *a priori* probability that the state  $X$  is  $x$ . If we receive two sets of measurements ( $Z_1, Z_2$ ) that are conditionally independent given the state  $X$ , i.e.,  $p(Z = Z_1, z_2 | X = x) = p(Z = z_1 | X = x)p(Z = z_2 | X = x)$ , then Equation 10 expands to

$$p(X = x | Z = z_1, z_2) \propto p(Z = z_1 | X = x)p(Z = z_2 | X = x)p(X = x). \quad (11)$$

Then, the MAP estimate of the posterior state  $\hat{x}$  is given as

$$\hat{x} = \arg \max p(Z = z_1 | X = x)p(Z = z_2 | X = x). \quad (12)$$

Assuming the likelihood function is modeled as a Gaussian distribution, i.e.,

$$p(Z = z | X = x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x - z)^2}{2\sigma^2}\right) \quad (13)$$

where the standard deviation  $\sigma$  denotes the uncertainty associated with the measurements provided by the sensor, then the MAP estimate of Equation 12 becomes

$$\hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z_2, \quad (14)$$

and the total variance of the fused measurements becomes  $\sigma_f^2 = (\sigma_1^{-2} + \sigma_2^{-2})^{-1}$ .

Essentially, Equation 14 describes the weighted mean of two measurements based on the uncertainties associated with each measurement sources. In the context of TRON calibration, the uncertainty associated with each measurement source is estimated assuming  $T_{TC,cal}$ , the true pose estimated from solving the PnP, represents the *mean* transformation. Then, the variance associated with each dimension of the estimated position vector  ${}^C\tilde{\mathbf{t}}_{CT}$  is defined as

$$\sigma_p^2 = \frac{1}{N-1} \sum_{i=1}^N \left[ ({}^C\tilde{\mathbf{t}}_{CT}^{(i)})_p - ({}^C\mathbf{t}_{CT,cal}^{(i)})_p \right]^2, \quad p \in \{1, 2, 3\} \quad (15)$$

so that they can be applied to Equation 14 to compute the dimension-wise weighted mean of two measurements.

To compute the weighted mean in the orientation space, the variance of measurement is taken as a scalar value corresponding to the angular distance between the measured and the mean rotation matrices, i.e.,

$$d\theta^{(i)} = \arccos\left(\frac{\text{tr}(d\mathbf{R}^{(i)}) - 1}{2}\right), \quad \text{where } d\mathbf{R}^{(i)} = (\mathbf{R}_{TC}^{(i)})^\top \mathbf{R}_{TC,cal}^{(i)}. \quad (16)$$

Then, the variance associated with the relative orientation is given as

$$\sigma_R^2 = \frac{1}{N-1} \sum_{i=1}^N (d\theta^{(i)})^2. \quad (17)$$

Given the uncertainty associated with each measurement, the weighted mean of rotation matrices from KUKA- and Vicon-based RWHE calibrations can be computed by solving

$$\tilde{\mathbf{R}}_{TC} = \arg \min_{\mathbf{R} \in SO(3)} \sum_{S \in \{K, V\}} w_S \|\mathbf{R} - \tilde{\mathbf{R}}_{TC,S}\|_F^2 \quad (18)$$

where  $w_K = \sigma_{R,V}^2 (\sigma_{R,K}^2 + \sigma_{R,V}^2)^{-1}$ ,  $w_V = \sigma_{R,K}^2 (\sigma_{R,K}^2 + \sigma_{R,V}^2)^{-1}$ . The solution to Equation 18 is given as<sup>37</sup>

$$\tilde{\mathbf{R}}_{TC} = \bar{\mathbf{R}} \mathbf{U} \mathbf{D}^{-1/2} \mathbf{U}^\top \quad (19)$$

where  $\bar{\mathbf{R}} = \sum_{S \in \{K, V\}} w_S \tilde{\mathbf{R}}_{TC,S}$ ,  $\mathbf{D}$  is the diagonal matrix consisting of the eigenvalues of  $\bar{\mathbf{R}}^\top \bar{\mathbf{R}}$ , and  $\mathbf{U}$  is an orthogonal matrix whose column vectors are the eigenvectors of  $\bar{\mathbf{R}}^\top \bar{\mathbf{R}}$  corresponding to the diagonal entries of  $\mathbf{D}$ .

### Rejection Criterion

As noted previously, it is necessary to develop a criterion based on which one chooses to either reject the Vicon measurement and opt for  $\tilde{\mathbf{T}}_{TC,K}$  or accept the Vicon measurement and perform data fusion. The rejection criterion is established based on the calibration samples which are collected in the optimal setting in terms of the IR marker visibility. Namely, the variance of  $\tilde{\mathbf{T}}_{TC,V}$  with respect to  $\tilde{\mathbf{T}}_{TC,K}$  is computed using an approach similar to Equations 15, 16, 17 during the calibration. Then, during the future data acquisition step, if the estimated pose  $\tilde{\mathbf{T}}_{TC,V}$  is  $1.96\sigma$  away (i.e., 95% confidence interval) from  $\tilde{\mathbf{T}}_{TC,K}$  in any of the position or orientation components, the Vicon measurement is deemed noisy and thus rejected.

## EXPERIMENT

This section first describes the experiment and result of the TRON calibration. Then, a trajectory with 111 pose samples is run to validate that the reported calibration performance extends to an arbitrary relative pose reconfigurable within the simulation room.

### Calibration

The calibration in this work uses the Point Grey Grasshopper 3 camera with a Xenoplan 1.4/17mm lens on a ceiling-mounted robot and a half-scale mockup model of the Tango spacecraft from the PRISMA mission<sup>20</sup> on the ground-fixed robot. As visualized in Figure 7, the calibration uses a 350 mm × 350 mm ChArUco board with  $11 \times 11$  pattern of 30 mm squares printed on a flat aluminum composite panel. The calibration first involves collecting  $N = 64$  samples from random orientations up to  $45^\circ$  tilt from a normal vector from the board. The separation between the camera and the board is kept around 0.75 m to ensure consistent accuracy of the detected pattern features.

The calibration results are reported as an accuracy of the poses estimated from the Vicon- and KUKA-only RWHE calibrations and the data fusion. These estimated poses are compared against those from PnP (i.e.,  $\mathbf{T}_{TC,cal}$ ) in terms of three metrics. The mean translation and orientation errors over  $N$  samples are reported as

$$E_T = \frac{1}{N} \sum_{i=1}^N \|{}^C \tilde{\mathbf{t}}_{CT}^{(i)} - {}^C \mathbf{t}_{CT,cal}^{(i)}\|, \quad (20)$$

$$E_R = \frac{1}{N} \sum_{i=1}^N \arccos \left( \frac{\text{tr}(\delta \mathbf{R}) - 1}{2} \right), \quad \delta \mathbf{R} = (\tilde{\mathbf{R}}_{TC})^\top \mathbf{R}_{TC,cal}. \quad (21)$$

**Table 1.** Results of TRON calibration based on single-source measurements and the data fusion technique. Mean values are reported along with one standard deviation errors.

Metrics	KUKA-only RWHE	Vicon-only RWHE	Data Fusion
$E_T$ [mm]	$2.429 \pm 0.866$	$1.208 \pm 0.678$	<b><math>0.815 \pm 0.494</math></b>
$E_R$ [ $^\circ$ ]	$0.637 \pm 0.120$	$0.172 \pm 0.083$	<b><math>0.169 \pm 0.077</math></b>
$E_p$ [pix]	$5.176 \pm 2.397$	$4.175 \pm 2.718$	<b><math>2.758 \pm 1.716</math></b>

The third metric is the mean of root-mean-square (RMS) reprojection error of the chessboard corners of the ChArUco board, which is reported as

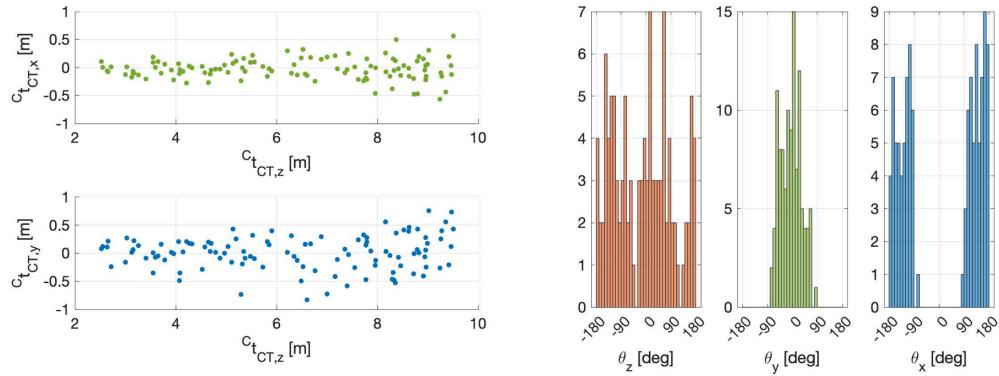
$$E_p = \frac{1}{N} \sum_{i=1}^N \sqrt{\frac{1}{M_i} \sum_{j=1}^{M_i} \|\Pi(\tilde{T}_{TC}^{(i)}, \mathbf{X}_j, \mathbf{K}, \mathbf{d}) - \mathbf{x}_j\|^2}. \quad (22)$$

The results of TRON calibration are shown in Table 1. First, the Vicon-only RWHE is a clear winner compared to KUKA-only RWHE in terms of all metrics, exhibiting millimeter-level position accuracy and millidegree-level orientation accuracy. The KUKA-only RWHE results in a much larger orientation error, which suggests that the fixtures between the end-effectors and the objects, especially the target model, are not perfectly rigid. This is most likely causing the target model to tilt with respect to its end-effector axis up to a couple of millidegrees. However, the data fusion shows improvement over both Vicon- and KUKA-only RWHE, reducing both mean translation and reprojection errors by nearly 33%. Despite the noticeable discrepancy in performances between KUKA- and Vicon-only RWHE, the performance of data fusion does not deteriorate because it performs a weighted average of  $\tilde{T}_{TC,V}$  and  $\tilde{T}_{TC,K}$ , with more weights assigned to the variable with less uncertainty. Since the Vicon measurements result in a better calibration performance, the estimated pose from the data fusion is largely close to  $\tilde{T}_{TC,V}$ , with a small correction by  $\tilde{T}_{TC,K}$  leading to improvement in all metrics.

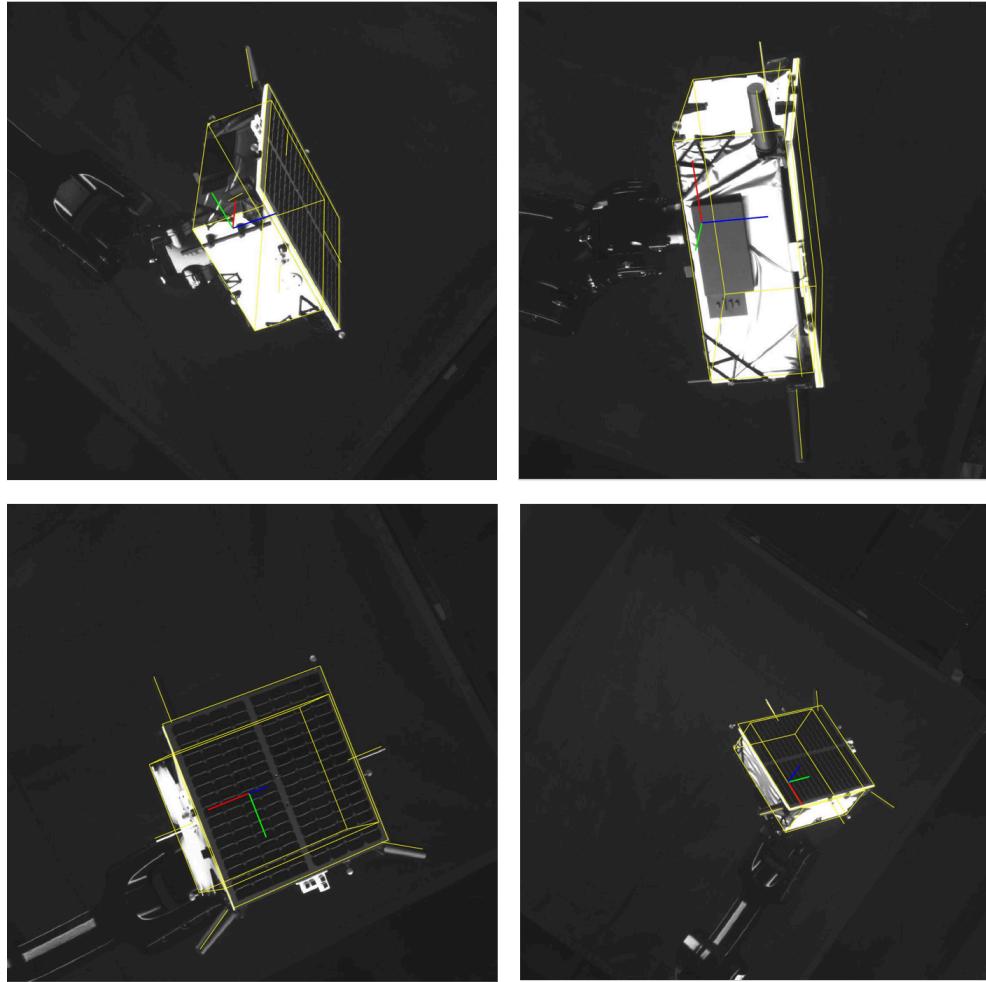
### Arbitrary Pose Reconfiguration

In order to visually confirm the validity of the calibration and data fusion results for an arbitrary pose reconfiguration, a separate dataset of 111 samples are collected based on the pose distribution visualized in Figure 8. Note that the relative position assumes a full-scale target, thus allowing a separation along the camera boresight up to nearly 9.5 meters. The select pose labels estimated via data fusion are visualized in Figure 9, where the Tango spacecraft’s full-scale wireframe model is projected onto the images based on the estimated poses. It shows that the calibration and the data fusion method result in consistent accuracy of the estimated pose labels regardless of the camera’s orientation and the distance to the target. In general, a similar trend is observed for most of the samples in the dataset to which the data fusion is applicable. It even extends to 15/111 samples in which the Vicon measurements are rejected.

However, it should be noted that there are multiple sources of errors that could spike the translation error up to a centimeter. One is that the model can tilt up to a couple of millidegrees, as evidenced from the calibration result shown in Table 1. Even such a small error could result in a non-negligible misalignment of the reprojected wireframe model when viewed at a close distance. This would likely happen in case the Vicon measurement, which results in a better orientation estimation, is rejected and must resort to a KUKA measurement. The other is any errors introduced



**Figure 8.** Position (left) and orientation (right) distributions of the test samples. The relative orientation is parametrized as Euler angles for visualization.



**Figure 9.** Visualization of estimated pose labels from data fusion via projection of the Tango wireframe model.

into the orientation offsets present in the camera-holding robot arm, i.e.,  $(\mathbf{T}_{C_K C}, \mathbf{T}_{C_V C})$ . In reality, if there is a non-negligible error in either of these orientation offsets that results in deviation of the camera boresight, the resulting error in the target's position would scale with the separation between the camera and the target. For example, if the camera boresight is tilted by  $0.1^\circ$ , the target's position would be off by 1.75 cm when placed at 10 meters away. In fact, such an error could be introduced in either KUKA or Vicon measurements as well. While no samples in the collected dataset exhibit such an egregious level of misalignment that could indicate a centimeter-level error, it still remains a possibility.

## COMPARATIVE ASSESSMENT OF IMAGE QUALITY

This section focuses on the assessment of the quality of images that can be created from the TRON testbed. First, a few select samples with various illumination conditions are showcased to visually demonstrate the range of possible lighting configurations that TRON can simulate using its light boxes and sun lamp. Then, their synthetic counterparts are rendered using the estimated pose labels of the simulated images for a comparative assessment of the domain gap between two imageries.

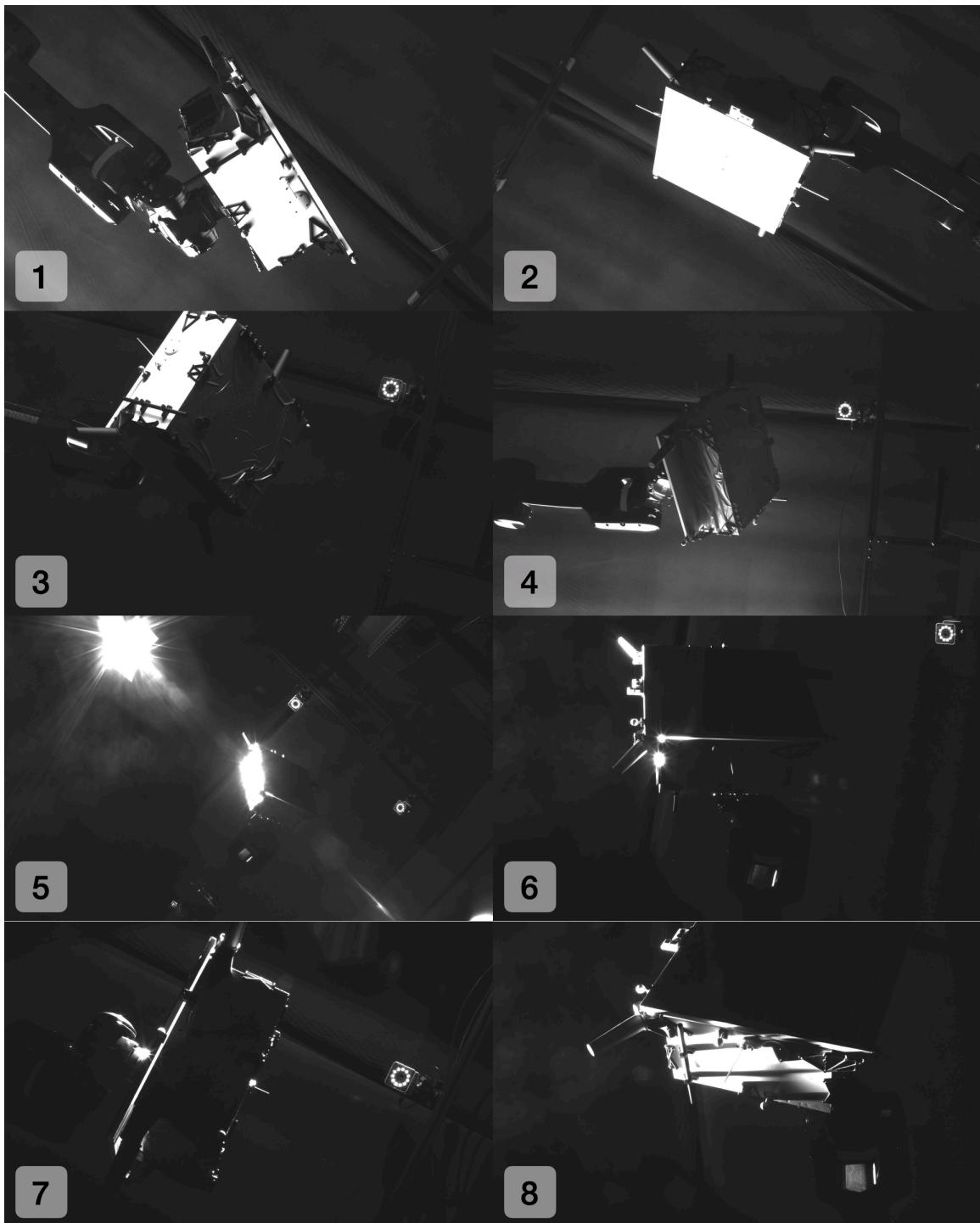
First, Figure 10 visualizes 8 simulated images with different pose and illumination configurations. The first 4 images (no. 1-4) show the effects of the Earth albedo boxes on the mockup Tango model. As expected from the albedo light, the model is quite evenly illuminated on the parts that face the activated light boxes. One peculiar feature is shown in image no. 2, in which the solar panel completely reflects the light from the albedo boxes it faces due to its high reflectivity. Such effect is not observed in the SPEED synthetic imagery.

The next 4 images (no. 5-8), on the other hand, show the effects of directed light from the sun lamp cast on the satellite model. Image no. 5 shows some of the boundaries between the model parts disappearing due to extensive directed light and the model surface's reflectivity. It also shows the sun lamp casting a flare effect across the image, as it often happens when the camera is directed near the Sun or any source of direct light. Images no. 6, 7 illustrate the high contrast within the model due to direct sunlight, the feature commonly encountered in spaceborne imageries. Finally, image no. 8 shows the shadow cast by individual parts of the spacecraft, the effect also unobserved in the SPEED synthetic imagery.

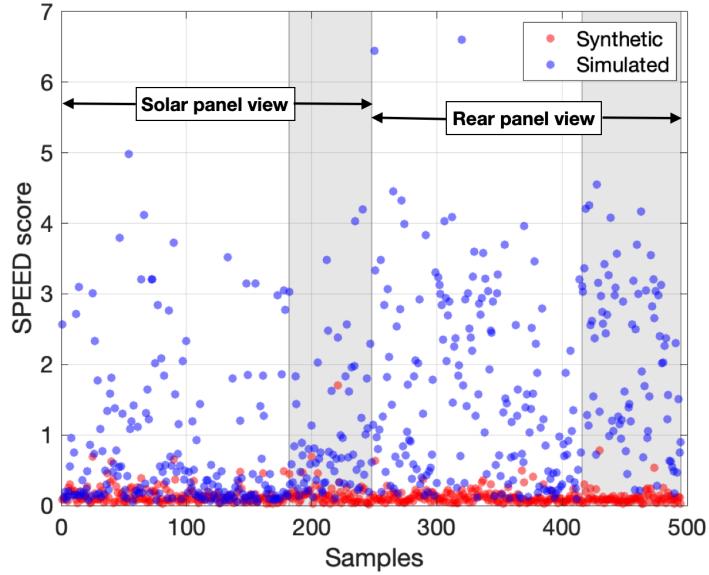
In order to quantitatively evaluate the domain gap between the TRON simulated images of Figure 10 and its synthetic counterparts rendered from OpenGL, a CNN by Park et al.<sup>14</sup> is pre-trained on SPEED synthetic training set and used as a reference model. Specifically, 495 simulated images are collected based on pose labels sampled from a full orientation space and a position distribution with minimum separation at 6.5 meters in order to align with the distribution of the SPEED synthetic training set.<sup>13,19</sup> Then, the synthetic images with the same pose labels are rendered with the same setting of the SPEED synthetic imagery and the direction of the light source approximately aligning with the location of the activated albedo boxes and the sun lamp. In order to minimize the distraction from any items in the background, both synthetic and simulated images are masked around the satellite. The CNN performance is measured by the SPEED score defined as

$$\text{SPEED Score} = \frac{1}{N} \sum_{i=1}^N E_R^{(i)} + E_T^{(i)} / \|{}^C \mathbf{t}_{CT}^{(i)}\| \quad (23)$$

where  $E_R^{(i)}$  is the rotation error in radians, and  $E_T^{(i)}$  is the translation error in meters. Essentially,



**Figure 10.** Representative illumination configurations that can be recreated in TRON. Images no. 1-4 use the Earth albedo light boxes, and images no. 5-8 use the sun lamp.



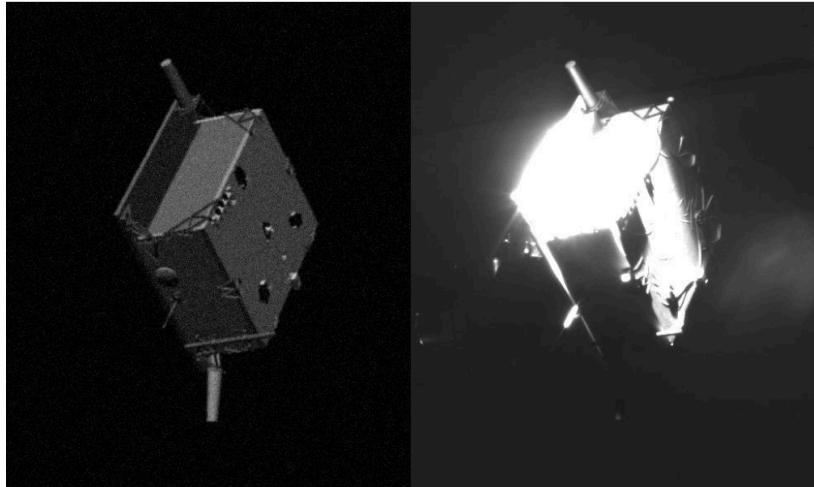
**Figure 11.** SPEED score of the pretrained CNN by Park et al.<sup>14</sup> on synthetic and simulated images of identical pose labels. The first half of the samples have solar panel views, while the other half have rear panel views. The data samples in the shaded region indicate the simulated images are illuminated by the sun lamp. The remaining samples are illuminated by the light boxes.

**Table 2. Mean SPEED scores for different configurations of the model view and light source.**

Configuration (view, light)	Synthetic	Simulated
Solar panel, light boxes	0.140	0.810
Solar panel, sun lamp	0.170	1.007
Rear panel, light boxes	0.121	1.568
Rear panel, sun lamp	0.114	2.062

SPEED score is the average of the sum of the rotation error and the translation error normalized by the norm of the ground-truth translation vector.

The CNN performance is visualized for all 495 pairs of the synthetic and simulated images in Figure 11. It shows in that in general, the SPEED score is worse for simulated images given the same pose labels and the directions of the light source. Table 2 then reports the average SPEED score for different categories of the 495 images. Namely, the collected images have the model mounted in either rear panel or solar panel side, which means the camera has the unobstructed view of the solar panel and rear panel, respectively. The images can also be illuminated by either the light boxes or the sun lamp. The combination of these categories results in 4 different configurations, and the distinctions are made in Figure 11 as well. Table 2 shows the CNN pre-trained on the SPEED synthetic training set reports on average far worse performance when using the sun lamp and viewing the rear panel of the Tango model. This means not only is the sun lamp casting a more challenging illumination effect, but also the rear panel of the mockup model exhibits a worse gap in the surface texture compared to the synthetic images. Such difference is visualized in Figure 12, where the syn-



**Figure 12. Comparison of synthetic vs. simulated images for rear panel view direction and the sun lamp illumination.**

thetic image lacks the texture of the simulated images on its rear panel and retains all of its model features despite the intended illumination of the Sun. This study suggests that the simulated images can be successfully used to measure and validate the domain gap against the synthetic images, and that they are even more powerful when 1) the sun lamp is used for illumination, and 2) the more textured sides of the model are shown.

## CONCLUSION

This paper provides a systematic introduction of the next generation Robotic Testbed for Rendezvous and Optical Navigation (TRON) facility at Stanford’s Space Rendezvous Laboratory (SLAB), the first-of-a-kind robotic testbed for validating the Convolutional Neural Networks (CNN) or any vision-based navigation algorithms for spaceborne applications. TRON is capable of reconfiguring an arbitrary relative pose between a camera and a mockup model of the target Resident Space Object (RSO) using its two 6 degrees-of-freedom KUKA robot arms, so that one can efficiently generate an arbitrary number of target images with maximally diverse pose distribution. The multi-source calibration of TRON enables the estimation of the true pose between the camera and the target with a millimeter-level position and millidegree-level orientation accuracy. TRON also consists of 10 Earth albedo light boxes and a metal halide arc lamp to emulate diffuse and direct light sources commonly encountered in space missions. The analyses of TRON simulated images with various illumination settings indicate that TRON is capable of generating an imagery that can be used to rigorously validate the robustness of CNNs on an unknown domain of imagery different from the synthetic training images.

Future TRON can benefit from a rigid mounting of a target mockup model and further upgrade to the KUKA robot arms which would drive the position accuracy of each robot to less than half a millimeter. By improving the performance of the KUKA-only calibration, the data fusion can benefit from more equal contributions from both KUKA and Vicon measurements instead of current one-sided dominance from Vicon. This would also remove a rare scenario where a correct pose estimate from a Vicon measurement is rejected in favor of an incorrect pose estimate from a KUKA measurement based on the current rejection criterion.

One of the most important use cases of TRON is the creation of the next generation Spacecraft Pose Estimation Dataset (SPEED), named SPEED+, which will include nearly 10,000 simulated images of the Tango spacecraft with various high-fidelity spaceborne illumination settings and even more synthetic images primarily intended for training. SPEED+ will facilitate the validation of a CNN trained to be robust on an unknown spaceborne domain without access to spaceborne images. Ultimately, SPEED+ will expand to include simulated images of different types of RSOs, such as satellites, space debris, asteroids, etc., so that it can help validate any vision-based navigation algorithms developed for future on-orbit servicing and space situational awareness missions.

## ACKNOWLEDGEMENT

The construction of the testbed was partly funded by the U.S. Air Force Office of Scientific Research (AFOSR) through the Defense University Research Instrumentation Program (DURIP) contract FA9550-18-1-0492, titled High-Fidelity Verification and Validation of Spaceborne Vision-Based Navigation. The authors would like to thank OHB Sweden for the 3D model of the Tango spacecraft used to create the images used in this article.

## REFERENCES

- [1] C. Harris and M. Stephens, “A combined corner and edge detector,” *In Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151.
- [2] J. F. Canny, “A computational approach to edge detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 8, June 1986, pp. 679–698, 10.1109/TPAMI.1986.4767851.
- [3] D. Ballard, “Generalizing the Hough transform to detect arbitrary shapes,” *Pattern Recognition*, Vol. 13, No. 2, 1981, p. 111–122, 10.1016/0031-3203(81)90009-1.
- [4] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, Vol. 60, No. 2, 2004, p. 91–110, 10.1023/b:visi.0000029664.99615.94.
- [5] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-Up Robust Features (SURF),” *Comput. Vis. Image Underst.*, Vol. 110, June 2008, pp. 346–359, 10.1016/j.cviu.2007.09.014.
- [6] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An Efficient Alternative to SIFT or SURF,” *Proceedings of the 2011 International Conference on Computer Vision, ICCV ’11*, Washington, DC, USA, IEEE Computer Society, 2011, pp. 2564–2571, 10.1109/ICCV.2011.6126544.
- [7] K. Dennison and S. D’Amico, “Comparing Optical Tracking Techniques in Distributed Asteroid Orbiter Missions Using Ray-Tracing,” *2021 AAS/AIAA Space Flight Mechanics Meeting, Virtual Event*, February 1-4 2021.
- [8] S. D’Amico, M. Benn, and J. L. Jørgensen, “Pose estimation of an uncooperative spacecraft from actual space imagery,” *International Journal of Space Science and Engineering*, Vol. 2, No. 2, 2014, p. 171, 10.1504/ijspacese.2014.060600.
- [9] S. Sharma, J. Ventura, and S. D’Amico, “Robust Model-Based Monocular Pose Initialization for Noncooperative Spacecraft Rendezvous,” *Journal of Spacecraft and Rockets*, 2018, p. 1–16, 10.2514/1.a34124.
- [10] V. Capuano, S. R. Alimo, A. Q. Ho, and S.-J. Chung, “Robust Features Extraction for On-board Monocular-based Spacecraft Pose Acquisition,” *AIAA Scitech 2019 Forum*, 10.2514/6.2019-2005.
- [11] V. Capuano, K. Kim, J. Hu, A. Harvard, and S.-J. Chung, “Monocular-based pose determination of uncooperative known and unknown space objects,” *69th International Astronautical Congress (IAC), Bremen, Germany*, 1-5 October 2018.
- [12] S. Sharma and S. D’Amico, “Reduced-Dynamics Pose Estimation for Non-Cooperative Spacecraft Rendezvous using Monocular Vision,” *40th Annual AAS Guidance and Control Conference, Breckenridge, Colorado*, February 2-8 2017.
- [13] S. Sharma and S. D’Amico, “Pose Estimation for Non-Cooperative Spacecraft Rendezvous Using Neural Networks,” *2019 AAS/AIAA Space Flight Mechanics Meeting, Ka’apalani, Maui, HI*, January 13-17 2019.
- [14] T. H. Park, S. Sharma, and S. D’Amico, “Towards Robust Learning-Based Pose Estimation of Noncooperative Spacecraft,” *2019 AAS/AIAA Astrodynamics Specialist Conference, Portland, Maine*, August 11-15 2019.

- [15] M. Kisantal, S. Sharma, T. H. Park, D. Izzo, M. Märkens, and S. D'Amico, "Satellite Pose Estimation Challenge: Dataset, Competition Design and Results," *IEEE Transactions on Aerospace and Electronic Systems*, 2020, pp. 1–1.
- [16] L. P. Cassinis, R. Fonod, E. Gill, I. Ahrns, and J. G. Fernandez, *CNN-Based Pose Estimation System for Close-Proximity Operations Around Uncooperative Spacecraft*, 10.2514/6.2020-1457.
- [17] B. Chen, J. Cao, Á. P. Bustos, and T.-J. Chin, "Satellite Pose Estimation with Deep Landmark Regression and Nonlinear Pose Refinement," *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 2816–2824.
- [18] S. Sharma, *Pose Estimation of Uncooperative Spacecraft using Monocular Vision and Deep Learning*. PhD thesis, Stanford University, Department of Aeronautics & Astronautics, Aug 2019.
- [19] S. Sharma, T. H. Park, and S. D'Amico, "Spacecraft Pose Estimation Dataset (SPEED)," Stanford Digital Repository. Available at: <https://doi.org/10.25740/dz692fn7184>, 2019.
- [20] S. D'Amico, P. Bodin, M. Delpech, and R. Noteborn, "PRISMA," *Distributed Space Missions for Earth System Monitoring Space Technology Library* (M. D'Errico, ed.), Vol. 31, ch. 21, pp. 599–637, 2013, 10.1007/978-1-4614-4541-8\_21.
- [21] P. Tsiotras, "ASTROS: A 5DOF experimental facility for research in space proximity operations," *Advances in the Astronautical Sciences*, Vol. 151, 01 2014, pp. 717–730.
- [22] M. Dor and P. Tsiotras, "ORB-SLAM Applied to Spacecraft Non-Cooperative Rendezvous," *2018 Space Flight Mechanics Meeting*, 10.2514/6.2018-1963.
- [23] R. Zappulla, J. Virgili-Llop, C. Zagaris, H. Park, and M. Romano, "Dynamic Air-Bearing Hardware-in-the-Loop Testbed to Experimentally Evaluate Autonomous Spacecraft Proximity Maneuvers," *Journal of Spacecraft and Rockets*, Vol. 54, No. 4, 2017, pp. 825–839, 10.2514/1.A33769.
- [24] Y. K. Nakka, R. C. Foust, E. S. Lupu, D. B. Elliott, I. S. Crowell, S.-J. Chung, and F. Y. Hadaegh, "A Six Degree-of-Freedom Spacecraft Dynamics Simulator for Formation Control Research," *2018 AAS/AIAA Astrodynamics Specialist Conference, Snowbird, UT*, August 19-23 2018.
- [25] L. Pasqualetto Cassinis, A. Menicucci, E. Gill, I. Ahrns, and J. Gil Fernandez, "On-Ground Validation of a CNN-based Monocular Pose Estimation System for Uncooperative Spacecraft," *8th European Conference on Space Debris*, Vol. 8, 2021.
- [26] "KR 10 R1100 sixx C-WP," [https://www.kuka.com/-/media/kuka-downloads/imported/6b77eecacfe542d3b736af377562ecaa/0000235181\\_en.pdf](https://www.kuka.com/-/media/kuka-downloads/imported/6b77eecacfe542d3b736af377562ecaa/0000235181_en.pdf). Accessed April 29, 2021.
- [27] C. Beierle and S. D'Amico, "Variable-Magnification Optical Stimulator for Training and Validation of Spaceborne Vision-Based Navigation," *Journal of Spacecraft and Rockets*, Vol. 56, Feb 2019, pp. 1–13, 10.2514/1.A34337.
- [28] T. Boge, T. Rupp, K. Landzettel, T. Wimmer, C. Mietner, J. Bosse, and B. Thaler, "Hardware in the Loop Simulator für Renezvous und Docking Manöver," *Deutsche Luft- und Raumfahrtkongress 2009*, September 2009.
- [29] A. Tabb and K. M. A. Yousef, "Solving the robot-world hand-eye (s) calibration problem with iterative methods," *Machine Vision and Applications*, Vol. 28, No. 5-6, 2017, pp. 569–590.
- [30] "Vero: Compact Super Wide Camera by Vicon," <https://www.vicon.com/hardware/cameras/vero/>. Accessed April 29, 2021.
- [31] "Verification of Light-box Devices for Earth Albedo Simulation," Technical Note, Stanford Space Rendezvous Lab (SLAB), January (2016).
- [32] M. Shah, "Solving the Robot-World/Hand-Eye Calibration Problem Using the Kronecker Product," *Journal of Mechanisms and Robotics*, Vol. 5, 06 2013, 10.1115/1.4024473.
- [33] L. Aiguo, L. Wang, and D. Wu, "Simultaneous robot-world and hand-eye calibration using dual-quaternions and Kronecker product," *International Journal of the Physical Sciences*, Vol. 5, 09 2010.
- [34] S. Garrido-Jurado, R. M. n. Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic Generation and Detection of Highly Reliable Fiducial Markers under Occlusion," *Pattern Recogn.*, Vol. 47, June 2014, p. 2280–2292, 10.1016/j.patcog.2014.01.005.
- [35] 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, 2008, p. 155–166, 10.1007/s11263-008-0152-6.
- [36] M. Kumar, D. Garg, and R. Zachery, "A generalized approach for inconsistency detection in data fusion from multiple sensors," *2006 American Control Conference*, 2006, pp. 6 pp.–, 10.1109/ACC.2006.1656526.
- [37] M. Moakher, "Means and Averaging in the Group of Rotations," *SIAM J. Matrix Anal. Appl.*, Vol. 24, Jan. 2002, p. 1–16, 10.1137/S0895479801383877.