(Preprint) AAS 23-369

# VISION-BASED 3D RECONSTRUCTION FOR NAVIGATION AND CHARACTERIZATION OF UNKNOWN, SPACE-BORNE TARGETS

Kaitlin Dennison* and Simone D'Amico†

Vision-based 3D reconstruction (V3DR) is a low-power, low-cost, and readily available alternative to stereophotoclinometry and LiDAR for 3D shape recovery of an arbitrary target. In the context of this work, V3DR comprises stereovision (SV) and structure from motion (SfM) using sequential or synchronized images from any number of views. It can be utilized in a variety of space rendezvous objectives from target pose estimation to full simultaneous navigation and characterization. However, it is rarely employed in practice because implementation can be complicated due to interdependencies between 3D reconstruction and various mission design parameters. The complexity is compounded by three gaps in literature. 1) V3DR's sensitivity to state estimation, number of observers, target motion, and clock synchronization. 2) The nuance of V3DR using sequential images versus synchronized images from multiple agents. 3) How such sensitivities and the techniques used for V3DR generalize to targets of any size and composition. To overcome these three gaps and increase V3DR usability, this paper performs a comprehensive assessment of the sensitivity of V3DR with respect to several design parameters and two vastly different targets: a large asteroid and a small spacecraft. The results show that absolute position error has the greatest influence on SV performance, increasing the number of observers does not necessarily reduce SfM error, and multi-agent SV is robust to image desynchronization. The lessons learned from these results are showcased in a demonstration of characterization and pose estimation of a non-cooperative spacecraft as well as simultaneous navigation and characterization of an asteroid. These two case studies show a median SfM error of 19.07mm and a median SV error of 103.793m for the spacecraft and asteroid case studies, respectively.

## INTRODUCTION

Spacecraft rendezvous with an arbitrary target has applications from celestial body characterization[1–3] to orbital debris removal and spacecraft maintenance.[4–6] Knowledge of the target's rigid shape is an essential component of target navigation and relative pose estimation. This is typically accomplished with stereophotoclinometery (SPC) for celestial bodies, sometimes with the aid of LiDAR.[2,7,8] If the target is man-made, missions either assume the target's shape model is known a priori or use LiDAR to obtain it.[6,9] While these methods are highly accurate, they rely on assumptions or high size, weight, power, and cost (SWaP-C) technology. In contrast, vision-based 3D reconstruction (V3DR), which comprises stereovision (SV) and structure from motion (SfM) in the context of this work, capitalizes on equipment already ubiquitous across spacecraft to perform online shape recovery: monocular cameras. Thus, V3DR can be a low SWAP-C alternative to SPC

*Doctoral Candidate, Stanford University Aeronautics & Astronautics Dept., Durand Building, 496 Lomita Mall Stanford, CA 94305.

†Associate Professor, Stanford University Aeronautics & Astronautics Dept., Durand Building, 496 Lomita Mall Stanford, CA 94305.

and LiDAR for recovering a 3D model of a target, though to date, it has not been widely adopted due to implementation complexities.

SPC is a computation- and time-intensive method of recovering a high-fidelity shape and albedo model of a target,[7] especially of small celestial bodies. It requires Earth communication and lengthy observation periods, which inhibits full mission autonomy until the shape model is complete.[2,8] Additionally, such a high-fidelity model is not strictly necessary for performing simultaneous navigation and characterization (SNAC) of celestial bodies.[1] In contrast, missions to rendezvous with a man-made object often rely on an a priori model of the target.[6,9] However, no model would typically be available for instances of space domain awareness or orbital debris. Furthermore, if there is a model, it could be inaccurate due to spacecraft damage.[6]

LiDAR is a general-use sensor that recovers the depth to a surface without any assumed knowledge. While LiDAR has high precision, it comes with high power requirements and demands trade-offs between moving parts, measurement spread, and observability range.[6,10] V3DR, on the other hand, can be performed onboard spacecraft, has low power requirements, and has high measurement spread and observability range at the expense of measurement precision and lighting invariance.[11–13] Monocular cameras are light enough to use on nanosatellites and provide numerous uses other than depth recovery.[8,14] Furthermore, the use of multiple agents in a distributed space system (DSS) improves overall mission flexibility and observability.[1,14] V3DR is a common component of SNAC[1] and its subset, simultaneous localization and mapping[15,16] (SLAM).

There are studies on V3DR (and highly-related keypoint matching) performance with respect to lighting,[17] keypoint detection and matching,[17,18] observer angular separation,[19] and mathematical techniques.[13,20] Nevertheless, there are three gaps in the space-rendezvous domain literature that make V3DR difficult to use in practice. One, sensitivity to observer state knowledge, target motion, and clock synchronization is especially lacking. Two, sequential imaging using a single camera has not been compared to simultaneous imaging using multiple synchronized cameras for moving, space-borne targets. Three, space rendezvous with respect to poorly-known asteroids and man-made objects are typically siloed in literature, preventing the harmonization of associated techniques.

As a result of these three knowledge gaps, many implementations of V3DR rely on iterative practices to identify the optimal initial image pairing, requiring many initial images across a variety of views.[12,21] They also sometimes rely on manually adjusting design parameters (e.g. initial image pair or feature detection), leading to a mission that is not truly autonomous.[2,8] If the space rendezvous field had a deeper understanding of how V3DR performance relates to many mission design parameters, both iteration and manual tuning could be eliminated.

This paper performs an assessment of V3DR performance with respect to various mission design parameters to fill the three gaps in literature and create a more complete picture of how to optimize a space rendezvous mission for fast, low SWaP-C structure recovery of an arbitrary target. Recommendations for mission design are then extrapolated from the simulation results. The remainder of this paper is organized as follows. First, the V3DR methods are explained. Next, the design parameter simulations are set up with each parameter model explicitly detailed. Then the design parameter simulation results are presented and discussed. Finally, two case studies are designed and evaluated based on the recommended V3DR usage from the simulations.

## VISION-BASED 3D RECONSTRUCTION METHODS

This paper focuses on performing SV* or SfM† on sparse features on the surface of the target using sequential or synchronized images from any number of views. SV and SfM are the basis of many other V3DR techniques so their characteristics can be generalized. The specific techniques for SV and SfM are not the subject of this work, but their performance with respect to various parameters is.

Both SV and SfM assume 2D points in the images have been associated with one another. Once matched, the 3D position of each set of 2D points can be computed via SV or SfM, depending on the level of assumed camera state knowledge. SV assumes the observer pose (position and attitude) is known with respect to the target's body-fixed (TBF) frame; SfM does not make that assumption and recovers the observer pose along with the 3D points. Frequently in V3DR problems, points are constructed in a single camera's frame of reference, called the primary observer (PO) frame.[11,15] The choice of PO is explained in the Simulation Setup section.

### Feature Detection and Matching

Two point types are used as measurements in this paper: shape model vertices and keypoint descriptors. Shape model vertices are the subset of 3D model vertices visible to the observer from the shape model used to generate the images. The vertices are projected into the image frame and provide an exact ground truth for matched features and recovered 3D points.

However, model vertices do not take perspective changes, lighting, and image noise into account so keypoints are also evaluated. SIFT[22] and ORB[23] are used for natural and man-made targets, respectively, as comparative assessments have shown each to work best for different scene types.[17,18,24] Both SIFT and ORB keypoints are first matched by their descriptors.[22] Such keypoint matching is often unreliable, so the epipolar constraint[1,11] is applied as an outlier rejection method using the true spacecraft pose information. Outliers are also rejected based on the Euclidean distance between their 3D ray-traced points. While these methods are unrealistic in practice, having near-perfect matching allows V3DR performance to be characterized independently from point association.

### Stereovision

SV uses the observer pose information to compute a least-squares estimate of the 3D positions of 2D points matched between images. The process of computing the 3D position of a single point in an arbitrary reference frame is described in Alg. 1. Looking at Alg. 1, $N$ is the total number of views, $\{\boldsymbol{l}\}_n$ is the set of 2D pixel measurements $\boldsymbol{l}$ that correspond to the features detected in the $n$th image, $\hat{\boldsymbol{l}}$ is the normalized form of $\boldsymbol{l}$. Lines 1-3 iterate over the views to normalize $\{\boldsymbol{l}\}_n$. This normalization procedure is the one performed in the normalized 8-point algorithm[11] to provide stability during singular value decomposition.

The for-loop starting on line 4 iterates over each set of pixel measurements $\{\boldsymbol{l}\}_i$ correlated amongst two or more views and computes an estimate of their corresponding 3D point $\tilde{\boldsymbol{L}}_i$. First, the corresponding $\hat{\boldsymbol{l}}$ values and observer pose information (3D positions $\boldsymbol{r}_n \in \mathbb{R}^3$ and attitude $\mathbf{R}_n : \mathbb{R}^3 \rightarrow \mathbb{R}^3$) are gathered into sets in lines 5-7. The choice of reference frame for $\boldsymbol{r}_n$ is arbitrary but all $\boldsymbol{r}_n$ must be in the same frame, $\mathbf{R}_n$ is a $3 \times 3$ rotation from the frame of $\boldsymbol{r}_n$ to the $n$th

*Stereovision using sparse features is sometimes referred to as triangulation.

†Structure from motion is sometimes considered a form of simultaneous localization and mapping (SLAM).

**Algorithm 1** Stereovision

1: **for** $n$ in $\{1, ..., N\}$ **do**
2: $\quad \{\hat{\boldsymbol{l}}\}_n \leftarrow$ normalize$(\{\boldsymbol{l}\}_n)$
3: **end for**
4: **for** each $\{\boldsymbol{l}\}_i$ across all $n$ **do**
5: $\quad \{\hat{\boldsymbol{l}}\}_i \leftarrow$ the normalized version of $\{\boldsymbol{l}\}_i$
6: $\quad \{\boldsymbol{r}_n\}_i \leftarrow$ the set of $\boldsymbol{r}_n$ from the views that observed $\{\boldsymbol{l}\}_i$
7: $\quad \{\mathbf{R}_n\}_i \leftarrow$ the set of $\mathbf{R}_n$ from of the views that observed $\{\boldsymbol{l}\}_i$
8: $\quad \check{\boldsymbol{L}}_i \leftarrow$ DLT$(\{\hat{\boldsymbol{l}}\}_i, \{\boldsymbol{r}_n\}_i, \{\mathbf{R}_n\}_i)$
9: $\quad \tilde{\boldsymbol{L}}_i \leftarrow$ refine $\check{\boldsymbol{L}}_i$ using $\{\boldsymbol{l}\}_i, \{\boldsymbol{r}_n\}_i,$ and $\{\mathbf{R}_n\}_i$
10: **end for**

camera frame, and $\tilde{\boldsymbol{L}}_i$ is computed in the same frame as $\boldsymbol{r}_n$. In line 8, an initial estimate $\check{\boldsymbol{L}}_i$ for $\tilde{\boldsymbol{L}}_i$ is computed using the direct linear transform (DLT).[11] Line 9 refines $\check{\boldsymbol{L}}_i$ to the final estimate by minimizing its reprojection error using $\{\boldsymbol{l}\}_i$ as described in Dennison et al. (2023).[1]

### Structure from Motion

SfM recovers the 3D points and the 6 degree of freedom poses of the observers up to an unknown scale. To accomplish this, the poses of all cameras and the 3D positions of all landmarks must be initialized with respect to the PO frame as described in Alg. 2. Alg. 2 is largely based on Nistér (2004)[25] and Enqvist et al. (2011);[20] it is similar to the *initial pose estimation via global relocalization* method in ORB-SLAM.[21]

**Algorithm 2** Structure from motion initialization

**for** each $\{\boldsymbol{l}\}_i$ across all $n$ **do**
$\quad \{\boldsymbol{l}\}_{i,1N} \leftarrow$ the set of measurements in $\{\boldsymbol{l}\}_i$ from views 1 and $N$ only.
**end for**
$\left\{{}^1\check{\boldsymbol{L}}\right\}_{1N}, {}^1\check{\boldsymbol{r}}_N, \underset{1\to N}{\check{\mathbf{R}}} \leftarrow$ Nistér's 5-point method using all $\{\boldsymbol{l}\}_{i,1N}$
**for** $n$ in $\{2, ..., (N-1)\}$ **do**
$\quad$ **for** each ${}^1\check{\boldsymbol{L}}_i$ in $\left\{{}^1\check{\boldsymbol{L}}\right\}_{1N}$ **do**
$\quad\quad \{\boldsymbol{l}\}_{i,1n} \leftarrow$ the set of $\boldsymbol{l}$ values from views 1 and $n$ in the $\{\boldsymbol{l}\}_i$ that corresponds to ${}^1\check{\boldsymbol{L}}_i$
$\quad$ **end for**
$\quad {}^1\check{\boldsymbol{r}}_n, \underset{1\to n}{\check{\mathbf{R}}} \leftarrow$ perspective-three-point solver using $\left\{{}^1\check{\boldsymbol{L}}\right\}_{1N}$ and all $\{\boldsymbol{l}\}_{i,1n}$
**end for**
$\left\{{}^1\check{\boldsymbol{L}}\right\} \leftarrow$ Alg. 1 using $\left\{{}^1\check{\boldsymbol{r}}_n\right\}, \left\{\underset{1\to n}{\check{\mathbf{R}}}\right\},$ and $\{\boldsymbol{l}\}$
$\left\{{}^1\tilde{\boldsymbol{L}}\right\}, \left\{{}^1\tilde{\boldsymbol{r}}_n\right\}, \left\{\underset{1\to n}{\tilde{\mathbf{R}}}\right\} \leftarrow$ bundle adjustment using $\left\{{}^1\check{\boldsymbol{L}}\right\}, \left\{{}^1\check{\boldsymbol{r}}_n\right\}, \left\{\underset{1\to n}{\check{\mathbf{R}}}\right\}$, and $\{\boldsymbol{l}\}$

Alg. 2 starts by iterating over the sets of correlated pixel measurements to extract the measurements $\{\boldsymbol{l}\}_{i,1N}$ that were taken by the outermost views, views 1 and $N$, only. All $\{\boldsymbol{l}\}_{i,1N}$ sets are used in Nistér's 5-point method[25] to compute an initial estimate for each 3D point ${}^1\check{\boldsymbol{L}}_i$ in the PO's camera frame, the position of the $N$th observer in the PO's camera frame ${}^1\check{\boldsymbol{r}}_N$, and the rotation matrix from the PO's camera frame to the $N$th observer's camera frame $\underset{1\to N}{\check{\mathbf{R}}}$. Note that some $\{\boldsymbol{l}\}_{i,1N}$ sets will be empty and Nistér's 5-point method may find some $\{\boldsymbol{l}\}_{i,1N}$ to be invalid. In these cases, no corresponding ${}^1\check{\boldsymbol{L}}_i$ is computed and $\left\{{}^1\check{\boldsymbol{L}}\right\}_{1N}$ consists of just the ${}^1\check{\boldsymbol{L}}_i$ values computed by Nistér's 5-point method. It is possible for the 5-point method to fail due to random sampling or an ambiguous similarity transform. These failures are not removed from any simulations in this paper.

The for-loop starting on line 5 estimates the position and attitude of the $n$th camera in the PO's camera frame using a perspective-three-point (P3P) solver.[26] To do so, lines 6-8 pull the pixel measurements that correspond to views 1 and $n$ from the original $\{\boldsymbol{l}\}_i$ set that the $\{\boldsymbol{l}\}_{i,1N}$ came from for the respective ${}^1\check{\boldsymbol{L}}_i$. In line 9, the set of estimated ${}^1\check{\boldsymbol{L}}_i$ values and the corresponding $\{\boldsymbol{l}\}_{i,1n}$ sets to compute the $n$th camera's position and attitude.

Line 11 uses the pose estimates and the original pixel measurements in Alg. 1 to compute an estimate for all correlated sets of measurements observed, not just those observed by both the PO and the $N$th observer. Finally, all poses and 3D feature position estimates are refined using bundle adjustment (BA).[27]

After this initialization, any subsequent camera poses are computed by first using the P3P solver to estimate the new pose, using Alg. 1 to compute all 3D feature position estimates, and then applying BA. Incremental SfM can be performed using BA on a sliding window of views as described in Mouragnon et al. (2009).[28] However, if the original SfM initialization is inaccurate, any additional camera pose computation will likely fail.[20] Thus, only initialization is analyzed in this paper.

## SIMULATION SETUP

In order to assess the performance of V3DR with respect to multiple mission design parameters, the simulation is set up such that each parameter can be easily modified and be as agnostic as possible to the target and camera. The parameters to be evaluated include the number of views, perspective change between views, distance to the target, observer pose uncertainty, clock synchronization, and target rotation. These are expanded upon throughout this section.

Two space environments are simulated: one with asteroid 433 Eros[1] as the target and the other with the Tango spacecraft from the PRISMA mission[29] as the target. Eros and Tango differ drastically in three areas. One, Eros' maximum diameter is 22,000 times Tango's width. This influences the orbit design and how the perspective changes. Two, Eros acts as the central body of the observers' orbits while Tango orbits the Earth along with the observers. Thus, the simulations must be designed with both absolute and relative orbital elements in mind. Three, the textures on the targets' surfaces are entirely different: Eros is a rocky body with natural features while Tango is a man-made object with harsh, geometric features. Generalizing between two radically different targets allows the lessons learned from the simulations to be used to navigate about and characterize space-borne targets of almost any size, orbit, and texture.

This section first sets up the relative geometry and dynamics of the simulations and explains how target generalization is achieved. Then it describes how one might design an actual swarm formation to match the simulation setup. Next, the image generation and processing pipeline is explained. Finally, the parameter evaluation simulations are explicitly defined.

### Relative Geometry

The parameters that influence the relative geometry are the number of views, perspective change between views, and the distance to the target. The relative geometry shown in Figure 1 is designed so each parameter can be independently modified. Observers are distributed in a circle as though they are each in a circular, relative orbit about the target. The central axis in Figure 1 is an arbitrary reference frame $F$ whose x- and z-axes are denoted with $X_F$ and $Z_F$, respectively. The y-axis $Y_F$ points into the page. This unconventional axis orientation was chosen to match the camera frame axis, which has its boresight down the positive z-axis. $F$ is aligned with the target's center but it does not rotate with the target, similar to a radial-transverse-normal frame.

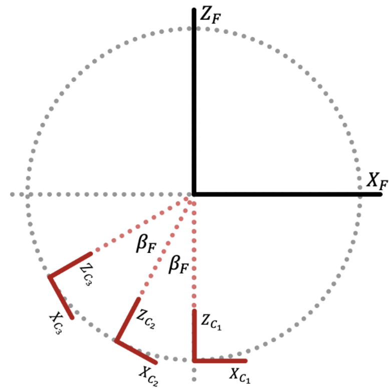



**Figure 1**: The relative geometry for 3 observers in the $F$ reference frame. Additional observers are added clockwise from the last one. Camera axes are denoted with $X_{C_n}$ and $Z_{C_n}$

The $n$th view, whether it is from the $n$th agent in a swarm or the $n$th image in a sequence taken from one agent, has a position ${}^F\boldsymbol{r}_n \in \mathbb{R}^3$ and attitude $\underset{F\to C_n}{\mathbf{R}} : \mathbb{R}^3 \to \mathbb{R}^3$ represented by

$$ {}^F\boldsymbol{r}_n = \mathbf{R}_2(-(n-1)\beta_F)[0,0,-d]^T, \quad \underset{F\to C_n}{\mathbf{R}} = \mathbf{R}_2((n-1)\beta_F), \tag{1} $$

where $\mathbf{R}_2$ is a passive y-axis rotation, $\beta_F$ is the angular separation between views in F, and $d$ is the distance to the target. The $n = 1$ observer is designated as the PO. The PO's camera frame is simply a translation of $F$ and is denoted by $X_{C_1}$ and $Z_{C_1}$. The camera boresight always points towards the target.

The distance to the target influences how much of field of view ($\theta_{\text{FOV}}$) the target takes up. With two different sizes of targets, the same $d$ and camera will result in a vast difference in perspective of the target. Thus, the percent of the $\theta_{\text{FOV}}$ that the target spans

$$ \gamma = 100 \times \frac{2}{\theta_{\text{FOV}}} \arctan \frac{D}{2(f-d)} \tag{2} $$

is evaluated instead of $d$ in the simulations. Here, $D$ is the maximum diameter of the target (33km for Eros and 1.48m for Tango) and $f$ is the camera focal length. This equation is derived from the Gaussian lens equation and the equation for the field of view.[30]

### Spacecraft Pose Uncertainty

There are two position error models used in this paper: absolute position error of each spacecraft and relative position error between each spacecraft and the PO. The former is representative of errors from state estimation systems like optical navigation or pseudorange and Doppler from the

Deep Space Network. The latter represents systems that use radio frequency links or GNSS signals where the inter-spacecraft range is measured with high accuracy compared to the absolute position of each spacecraft.[29,31]

Absolute position error is modeled as zero-mean Gaussian noise added to the $F$ position of each observer. The noisy position of the $n$th observer with absolute error applied is

$$ {}^{F}\breve{\boldsymbol{r}}_n = {}^{F}\boldsymbol{r}_n + \mathcal{N}(0, \sigma_a^2 \mathbf{I}_3), \tag{3} $$

where ${}^{F}\boldsymbol{r}_n$ is the true position of the observer in $F$ and $\sigma_a$ is the standard deviation of absolute position error, and $\mathbf{I}_3$ is a $3 \times 3$ identity matrix.

The relative position error is modeled as zero-mean Gaussian noise added to the difference between the $F$ position of $n$th observer and the PO. The noisy position of the $n$th observer with relative error applied is

$$ {}^{F}\breve{\boldsymbol{r}}_n = {}^{F}\breve{\boldsymbol{r}}_1 + {}^{F}\boldsymbol{r}_n - {}^{F}\boldsymbol{r}_1 + \mathcal{N}(0, \sigma_r^2 \mathbf{I}_3). \tag{4} $$

Here, $\sigma_r$ is the standard deviation of the relative position error and $\mathbf{I}_3$ is a $3 \times 3$ identity matrix. Nominal absolute position error is added to the PO when performing simulations with relative position measurements.

The attitude error is modeled as a 3-2-1 rotation respective to the roll, pitch, and yaw of the camera boresight. The noisy rotation matrix from $F$ to the $n$th camera frame $C_n$ is expressed as

$$ \underset{F \to C_n}{\breve{\mathbf{R}}} = \mathbf{R}_1\left(\mathcal{N}\left(0, (0.5\sigma_q)^2\right)\right) \mathbf{R}_2\left(\mathcal{N}\left(0, (0.5\sigma_q)^2\right)\right) \mathbf{R}_3\left(\mathcal{N}\left(0, \sigma_q^2\right)\right) \underset{F \to C_n}{\mathbf{R}}, \tag{5} $$

where $\mathbf{R}_1$, $\mathbf{R}_2$, and $\mathbf{R}_3$ are x-, y-, and z-axis rotations, respectively. The value $\sigma_q$ is the standard deviation of the attitude error. The roll has twice the error as the yaw and pitch, which is representative of pointing error for CubeSats.[32]

### Dynamics Model

Sequential imaging as well as the target rotation and clock synchronization parameters necessitate a dynamics model. A full model with perturbations and rigid body torques would make it difficult to modulate each parameter and isolate its effects. Instead, the observers simply travel along the circle in $F$ that they are distributed about. Each observer's angular velocity in $F$ is

$$ \omega = \sqrt{\mu / a^3}, \tag{6} $$

where $\mu$ is the gravitational parameter of the central body and $a$ is the semi-major axis of the orbit. If Eros is the target, it is the central body so $a = d$. If Tango is the target, it acts as the chief and the observers as deputies in a circular, relative orbit about Tango. The time to circumscribe the chief in a circular, relative orbit is the same as the chief's orbital period. Thus, $a = 8.413 \times 10^6$ m, which is approximately low Earth orbit.

In order to isolate the influence of target rotation rate $\dot{\theta}$, only single-axis target rotation is considered in this study. The target rotates about $Y_F$ so the angular separation $\beta$ between views in the TBF frame can be easily modulated depending on $\dot{\theta}$. Therefore, this simulation emulates an equatorial, circular orbit about the target.

For the single-agent case, $\beta_F = \omega \triangle t$ where $\triangle t$ is the time interval between images. Since $\beta$ is the parameter to be modulated, $\triangle t$ and $\beta_F$ are computed as

$$\triangle t = \frac{\beta}{|\omega - \dot{\theta}|}, \quad \beta_F = \frac{\omega\beta}{|\omega - \dot{\theta}|}. \tag{7}$$

If the target is stationary ($\dot{\theta} = 0$) with respect to $F$ then $\beta_F = \beta$. In the multi-agent case, the images are taken simultaneously so $\beta_F = \beta$ always. The influence of $\dot{\theta}$ is highly dependent on $\triangle t$ and $\beta$. It is likely that a particular $\triangle t$ is chosen for the entirety of the mission so when $\dot{\theta}$ is being evaluated, $\triangle t$ is kept constant, allowing $\beta$ to change. In all other evaluations, $\beta$ is set to a particular value.

Finally, the last component of dynamics is clock synchronization where the onboard clock drifts out of sync with GPS time. The drift rate is typically small enough that the clock error between sequential images from the same agent is inconsequential. However, every agent in a DSS has its own clock and these may become significantly out of sync with each other after long drift periods. Therefore, this paper only includes clock synchronization for multi-agent systems.

Clock synchronization error is usually assessed as part of the relative position measurement between two observers. For instance, pseudorange measurements $p_n$ between the PO and the $n$th agent can be modeled as

$$p_n = \| {}^F\boldsymbol{r}_n - {}^F\boldsymbol{r}_1 \|_2 + c\delta t_n - c\delta t_1, \tag{8}$$

where $c$ is the speed of light and $\delta t_n$ is the clock error between the PO and the $n$th agent.[31] However, Eq. (4) already assesses this in a more general form and there is another consequence of desynchronized clocks that specifically affects SV performance: image desynchronization.

In a DSS, all $N$ images are assumed to be acquired at the same predetermined time $t_{acq}$. However, if the clocks are not synchronized, each image is taken at a time $t_n = t_{acq} + \delta t_n$ (by definition, $\delta t_1 = 0$). The target and all of the observers will have a different pose at each $t_n$ than at $t_{acq}$, resulting in an erroneous SV computation because SV uses the observer and target poses at $t_{acq}$. Thus, in this sensitivity analysis, while the observer poses used for SV computation are defined using Eq. (1), the true scene used to generate each image is defined as follows. The desynchronized observer poses are

$${}^F\boldsymbol{r}_n^d = \mathbf{R}_2(-\omega\delta t_n)\, {}^F\boldsymbol{r}_n, \quad \underset{F\to C_n}{\mathbf{R}^d} = \mathbf{R}_2(-\omega\delta t_n) \underset{F\to C_n}{\mathbf{R}} \tag{9}$$

and the desynchronized target pose in the $n$th image is

$$\underset{TBF\to F}{\mathbf{R}^{d,n}} = \mathbf{R}_2(-\dot{\theta}\delta t_n) \underset{TBF\to F}{\mathbf{R}}. \tag{10}$$

The rotation matrix $\underset{TBF\to F}{\mathbf{R}}$ is the rotation from TBF to F. In a real mission, the SV error induced by image desynchronization will likely be compounded with the SV error induced by relative position error from clock drift and offset. The two are kept separate in this study so their individual contributions to SV error can be characterized.

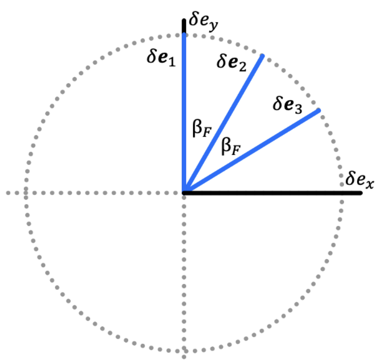
## Swarm Formation Examples

There are infinitely many ways to design a swarm formation and the simulations in this paper intend to provide guidelines for how one might modify an orbit design to improve V3DR performance. The relative geometry shown in Figure 1 is a basic swarm formation design intended for simulation purposes but can actually be achieved in practice.

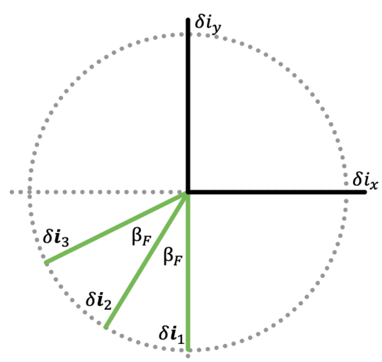
While full formation design is outside the scope of this paper, Table 1 presents example relative orbital elements (ROEs) for orbiting about each target. If Eros is the target, the PO is the chief and each subsequent observer is separated in the along-track direction. An inclined version of this formation design was used in Dennison et al. (2023).[1] If Tango is the target, Tango is the chief and the observers are separated so their $\delta\boldsymbol{e}$ and $\delta\boldsymbol{i}$ vectors are anti-parallel to achieve passive safety (as shown in Figures 2(a) and 2(b)). Note that the Tango-relative orbit may not result in the exact Figure 1 formation shown and will likely lead to a larger inter-observer angular separation than $\beta_F$.[33]

**Table 1**: The relative orbital elements of observer $n$ with respect to the primary observer ($n = 1$) when Eros is the target or with respect to Tango if Tango is the target. Here, $\tilde{\beta}(n) = (n-1)\beta_F$.

| **Target** | $\boldsymbol{a\delta a}$ | $\boldsymbol{a\delta\lambda}$ | $\boldsymbol{a\delta e_x}$ | $\boldsymbol{a\delta e_y}$ | $\boldsymbol{a\delta i_x}$ | $\boldsymbol{a\delta i_y}$ |
|---|---|---|---|---|---|---|
| Eros | 0 | $d\sin\tilde{\beta}(n)$ | 0 | 0 | 0 | 0 |
| Tango | 0 | 0 | $d\sin\tilde{\beta}(n)$ | $d\cos\tilde{\beta}(n)$ | $-d\sin\tilde{\beta}(n)$ | $-d\cos\tilde{\beta}(n)$ |




(a) Relative $\delta\boldsymbol{e}$ vectors.

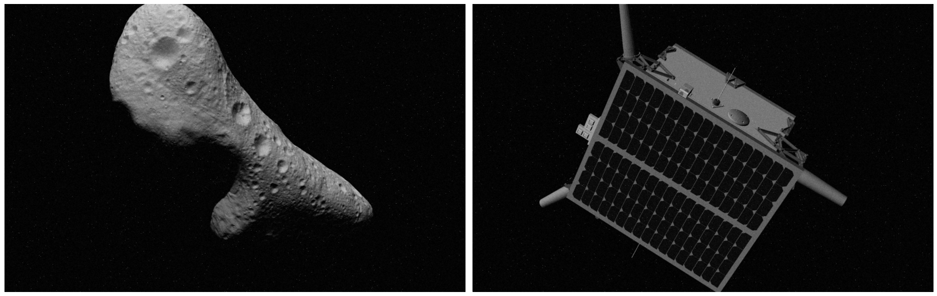



(b) Relative $\delta\boldsymbol{i}$ vectors.

**Figure 2**: The relative $\delta\boldsymbol{e}$ and $\delta\boldsymbol{i}$ vectors for the swarm formation with respect to Tango. Each $\delta\boldsymbol{e}_n$ and $\delta\boldsymbol{i}_n$ are antiparallel to achieve passive safety with a variable number of agents.

## Images

The Eros and Tango image-generation tools are those used in Dennison et al. (2023)[1] and Park et al. (2022).[4] Representative, synthetically generated images are shown in Figure 3. Each image is corrupted with a Gaussian white noise with a standard deviation of 0.0022, Gaussian blur of 0.8, and salt and pepper noise with a standard deviation of 0.001.[34,35] A median filter and a Wiener filter are applied to each image to aid keypoint detection and matching. The camera is modeled as a Grasshopper3[36] with $f = 17.5$mm, which is the same camera used in Park et al. (2022).[4]



(a) Asteroid 433 Eros (b) Tango

**Figure 3**: Examples of synthetically generated images with added noise.

**Design Parameters**

The tunable parameters explained throughout this section are compiled in the two following tables. Table 2 lists the parameters that define the relative geometry shown in Figure 1 as well as the dynamics model. The $\triangle t$ for $\dot{\theta}$ evaluations is computed using Eq. (7) and the default parameters. Table 3 lists the parameters that determine system uncertainties. The imaging clock error $\delta t_n$ is multiplied by the speed of light $c$ for readability. The nominal values in Table 3 are examples of realistic values for each parameter.[3,29,32]

**Table 2**: Observer-observer and observer-target relative geometry parameter ranges evaluated as well as default parameter settings. Eros's rotation rate[3] is $\dot{W} = 0.01897°/\text{s}$.

| | Angular Separation | Number of Views | Percent of FOV | Eros Rotation Rate | Tango Rotation Rate |
|---|---|---|---|---|---|
| Symbol | $\beta$ | $N$ | $\gamma$ | $\dot{\theta}$ | $\dot{\theta}$ |
| Units | ° | - | % | °/s | °/s |
| Minimum | 0.1 | 2 | 10 | -3$\dot{W}$ | -3 |
| Maximum | 90 | 7 | 150 | 3$\dot{W}$ | 3 |
| Default | 20 | 3 | 75 | $\dot{W}$ | 1 |

Randomized trials are used to assess V3DR performance with respect to each of the parameters in Tables 2 and 3. A simulation for each of the two different targets is run for each parameter. The parameter being evaluated is incremented $N_I$ times along its valid range. The parameters not being evaluated are set to their default values if they are in Table 2 or to zero if they are in Table 3.

There are $N_T$ trials performed for each increment with randomized target attitude and lighting conditions. The target's initial attitude in $F$ is a random quaternion vector. The lighting direction is a random unit vector pointing from the target center out into the hemisphere centered at the mean of all ${}^F\boldsymbol{r}_n$ vectors.

V3DR performance for a particular parameter iteration is measured as the median proportional depth recovery error $\delta\hat{\rho}$ across all 3D points recovered across all trials of that parameter increment. The proportional depth recovery error of a single point is

**Table 3**: Observer state uncertainty parameters evaluated. (E) and (T) denote Eros and Tango as the target, respectively. Nominal values are representative of past space missions.

| | **Absolute Position Error** | **Relative Position Error** | **Attitude Error** | **Imaging Clock Error** |
|---|---|---|---|---|
| Symbol | $\sigma_a$ | $\sigma_r$ | $\sigma_q$ | $c\delta t_n$ |
| Units | m | m | arcsec | m |
| Maximum (E) | 500 | 100 | 150 | $c(1\text{s})$ |
| Maximum (T) | 5 | 2 | 150 | $c(1\text{s})$ |
| Nominal (E) | 100 | 20 | 60 | 20 |
| Nominal (T) | 1 | 0.25 | 60 | 1 |

$$\delta\rho_i = \frac{\| {}^{C_1}L_i - {}^{C_1}L_{i*} \|_2}{D}. \tag{11}$$

Here, ${}^{C_1}L_i$ is the reconstructed point in the PO frame and ${}^{C_1}L_{i*}$ is the true (or ray-traced) point. Using the unitless $\delta\hat{\rho}$ abstracts the error from the scale of the target. The median absolute depth recovery error $\delta\hat{P}$, which has the same units as $D$, can be obtained from $\delta\hat{\rho}$ via $\delta\hat{P} = D\delta\hat{\rho}$.

## PARAMETER EVALUATION RESULTS & DISCUSSION

Each of the following subsections discusses the results of each design parameter simulation. For the relative geometry simulations, Figures 4-7 each contain a $2 \times 2$ grid of subplots with two y-axes: $\delta\hat{\rho}$ and average number of reconstructed points per trial $\hat{N}_p$ across all simulations in the plot (shown as the grey shaded region in the background). The rows are the SV-only (top) and SfM (bottom) simulation results. The columns are the simulation results using model vertices (left) and keypoint descriptors (right) as features. Similarly, Figures 8-10 contain the SV results for the observer state uncertainty simulations. They contain two subplots: the simulation results using model vertices (left) and keypoint descriptors (right) as features. All $\delta\hat{\rho}$ y-axes are logarithmic, and the $\delta\hat{\rho}$ data in Figures 4 and 6-10 are smoothed using a moving average. For context, a $\delta\hat{\rho} = 10^{-3}$ results in $\delta\hat{P}_{Eros} = 33$m and $\delta\hat{P}_{Tango} = 1.48$mm for Eros and Tango, respectively.

### Angular Separation

Looking at Figure 4, as $\beta$ increases, $\delta\hat{\rho}$ initially drops steeply and then plateaus. For all cases except SV using model vertices, $\delta\hat{\rho}$ slowly rises after the plateau. The average number of reconstructed points decreases as $\beta$ increases because matching diminishes as perspective changes. V3DR error and match counts decreasing as $\beta$ increases is a relationship already known in literature.[17,19,20]

The increase in $\delta\hat{\rho}$ as $\beta$ continues to increase is more interesting as it implies that increasing $\beta$ as much as possible can hinder performance. This relationship is most apparent for the keypoint descriptor simulations because of keypoints' weakness to large perspective changes. Furthermore, in Figure 4, SfM fails to compute a solution when $\beta$ gets too large because it does not have enough points. Therefore, there is an optimal range of values for $\beta$ that will depend on the prescribed

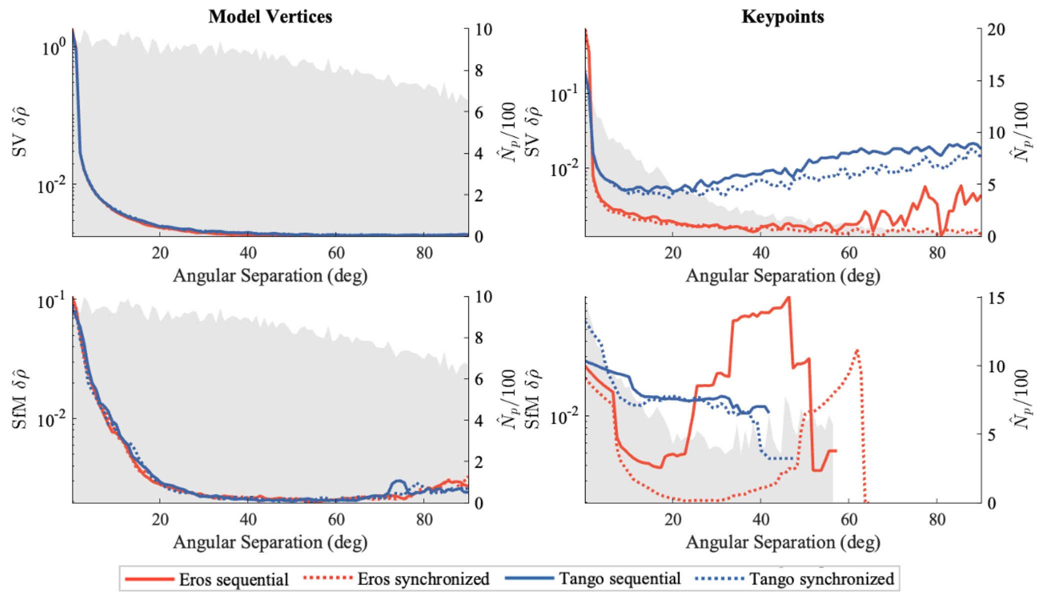



**Figure 4**: The effects of angular separation $\beta$ on V3DR performance.

relative geometry, feature descriptor, and V3DR method. E.g. for $N = 3$, a range of $\beta \in [10°, 60°]$ is acceptable for SV using SIFT but $\beta \in [15°, 40°]$ is more appropriate for SfM using ORB.

### Number of Views

In the left column of Figure 5, increasing $N$ decreases $\delta\hat{\rho}$. The change is highly correlated with $\beta$ because as $N$ increases, angular separation between the two outermost views increases. Unlike with $\beta$, increasing $N$ also increases $\hat{N}_p$, likely because there are more intermediate images to provide matches between features.

When the model vertices are swapped for keypoint descriptors, the relationship between $N$ and $\delta\hat{\rho}$ changes. While SV still loosely has a negative correlation between $N$ and $\delta\hat{\rho}$, SfM has a positive correlation. Spurious matches can still occur and increasing angular separation changes the perspective, diminishing match and keypoint centroid quality.

Increasing $N$ can benefit SfM for keypoints: $\hat{N}_p$ increases with $N$ and, consequently, the angular separation $\beta_{1,N}$ between the two outermost views can be increased further before failure than in the angular separation simulations. For instance, textitEros synchronized keypoints failed at $N = 6$ (or $\beta_{1,N} = 100°$) in the number of views simulation while *Eros synchronized keypoints* failed at $\beta = 64.53°$ in the angular separation simulation. Therefore, increasing $N$ can increase observability and the number of matches but it also can lead to poor matches if the views become too separated.

### Percent of FOV

In Figure 6, $\delta\hat{\rho}$ decreases as the target's percent of the FOV $\gamma$ increases. Because $\beta$ and $\dot{\theta}$ is constant, the viewing angle does not change with $\gamma$: only the camera's distance to the target is modulated. However, as $\gamma$ decreases, less detail is resolved and the 2D measurements become closer together in the image, leading to feature co-location in the presence of noise. This results

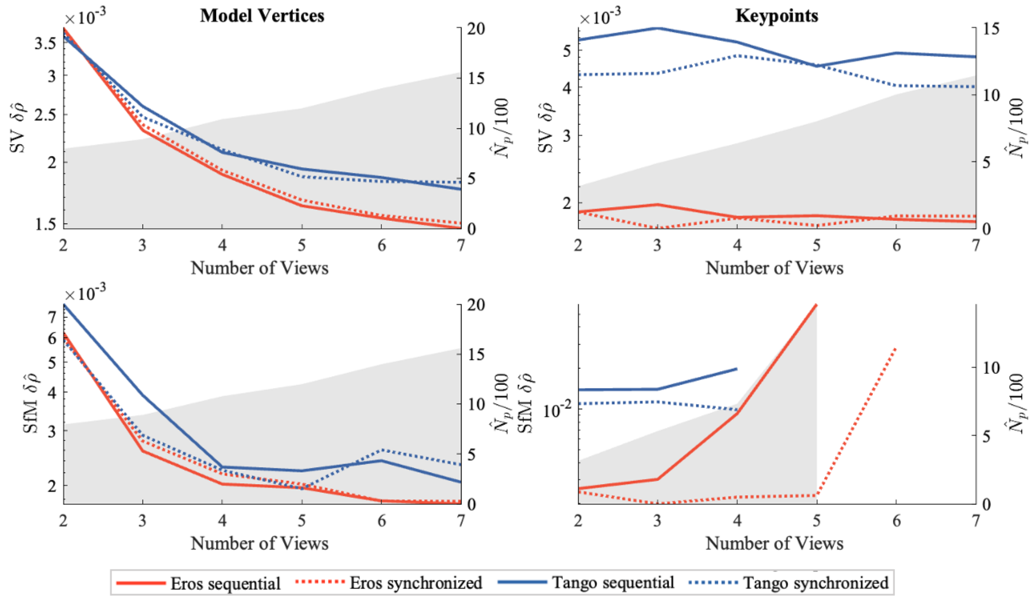



**Figure 5**: Simulation results on how additional views influence V3DR performance.

in poor numerical performance from V3DR,[19] increased influence of measurement noise, and less overall matches because matches are less certain.

The $\delta\hat{\rho}$ decrease becomes more gradual after approximately $\gamma = 40\%$ in all four plots. Additionally, $\hat{N}_p$ generally increases with $\gamma$ but the two keypoint-based plots have peaks in $\hat{N}_p$. Eq. (2) gives two ways of increasing $\gamma$: decreasing the distance to the target $d$ or modifying the camera parameters to decrease $f$ or $\theta_{\text{FOV}}$. Decreasing the distance to the target decreases mission safety but it is difficult to change camera parameters after launch.

### Rotation Rate

In Figure 7, it is clear that $\dot{\theta}$ influences sequential V3DR but not synchronized V3DR. This is because $\beta$ is related to $\dot{\theta}$ through Eq. (7) for sequential imaging only. For full loop-closure of systems like SNAC or SLAM where sets of synchronized images are taken sequentially,[1] V3DR will be influenced by $\dot{\theta}$.

There is a peak at $\dot{\theta} \approx \omega$ where the sequential images view the same scene every time and $\beta \approx 0$. Because the observer orbits the target's spin axis in this simulation, no change in the value of $\triangle t$ would be able to recover from this. This would not be the case for other orbit geometries (e.g. inclined with respect to the spin axis). Furthermore, $\delta\hat{\rho}$ increases as $|\dot{\theta}|$ increases when using keypoints because $\beta$ increases as well. So if $|\dot{\theta} - \omega| > 0$, an optimal $\triangle t$ and $\beta$ can be determined according to Eq. (7) and Figure 4.

### Absolute Position Error

Alg. 1 uses the absolute positions of the cameras to reconstruct the 3D points. Thus, the direct, positive correlation between the standard deviation of the absolute position error $\sigma_a$ and $\delta\hat{\rho}$ seen in Figure 8 is expected. Looking more closely, $\sigma_a$ has the greatest influence on $\delta\hat{\rho}$ of all the design parameters evaluated in this paper because it transfers error to $\delta\hat{\rho}$ at nearly $3\times$ magnification. For

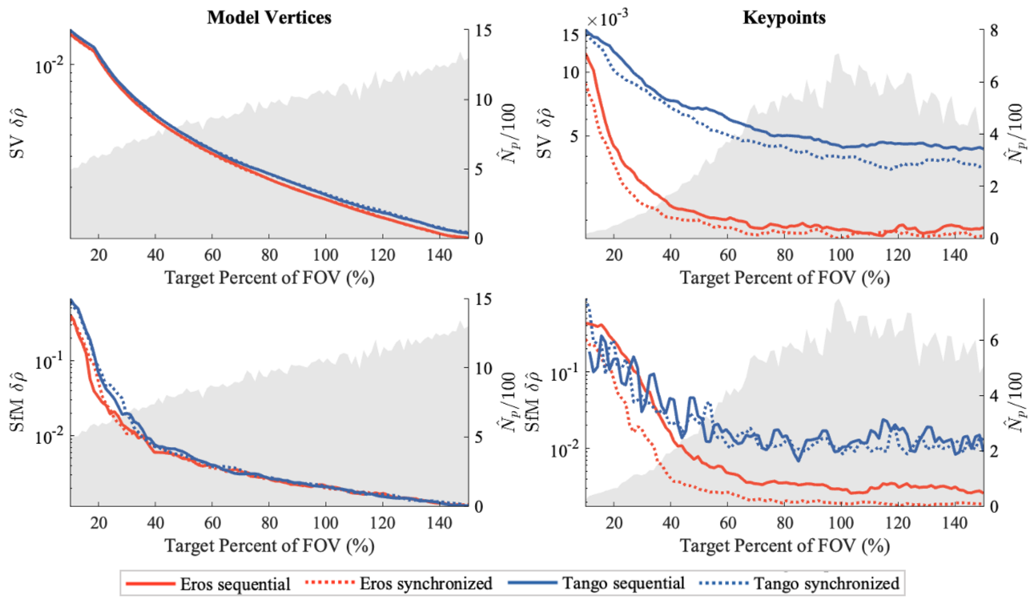



**Figure 6**: The effects of the percent of the field of view $\gamma$ the the target spans on V3DR performance.

instance, for Eros synchronized imaging using keypoints with nominal $\sigma_a$ ($\sigma_a = 100$m) the median error is $\delta\hat{\rho} = 0.011 \Rightarrow \delta\hat{P} = 363$m. The absolute position error comes from spacecraft navigation, whether it is terrain-relative or GPS-based. Minimizing $\sigma_a$ is a difficult task but one of the most important to mitigate 3D reconstruction error when using SV.

### Relative Position Error

According to Figure 9, using a relative position error allows a nearly order of magnitude reduction in SV error compared to all images having the same $\sigma_a$ applied. This is because the relative position error is an order of magnitude smaller than absolute position error. Furthermore, with nominal state uncertainties from Table 3, the SV error for Tango is still very high: $\delta\hat{\rho} = 0.843 \Rightarrow \delta\hat{P} = 1.24$m. The SV error for Eros comes to a reasonable level with relative position measurements: $\delta\hat{\rho} = 0.0025 \Rightarrow \delta\hat{P} = 82.5$m. This could be due to the large scale of the Eros system and not the keypoints themselves because the same trend is visible in the model vertices.

### Image Acquisition Clock Error

Figure 10 shows that image desynchronization has minimal impact on SV performance. In the these simulations, Tango has $\omega = 0.04687°$/s, $\dot{\theta} = 1°$/s, and $d = 3.01$m. If the clocks were off by 1s, the spacecraft's desynchronized position would be rotated by $0.0937°$ according to Eq. (9), adding $5.07$mm of error in the along-track direction to the SV computation. The $\delta\hat{\rho}$ for *Tango synchronized keypoints* in Figure 10 is $4.439 \times 10^{-3}$ ($\delta\hat{P} = 6.57$mm). This is slightly higher than $5.07$mm due to the rotation of Tango as expressed in Eq. (10) and the use of keypoints.

Comparatively, if pseudorange were used in this computation according to Eq. (8), the relative position error due to a clock error of 1s would be on the order of $3 \times 10^8$ m. Therefore, when the clocks are not perfectly synchronized, error will likely be propagated through the relative position error and not through the actual images becoming desynchronized.

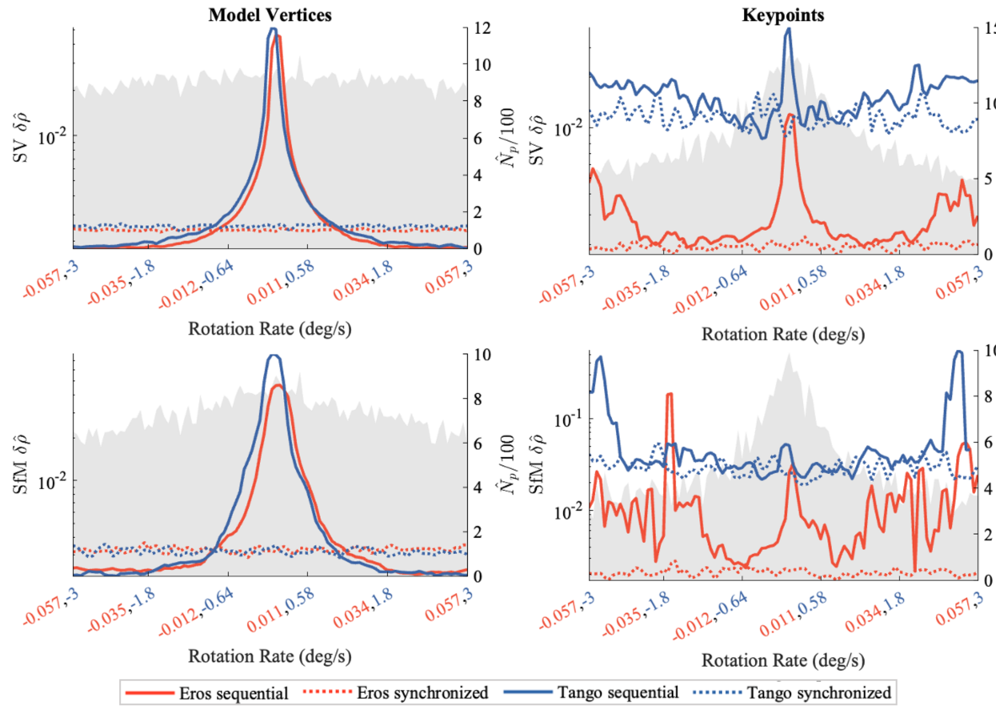



**Figure 7**: The effects of the target's rotation rate $\dot{\theta}$ on V3DR with a constant imaging rate $\triangle t$. The red and blue x-labels correspond to Eros and Tango, respectively.

### Attitude Error

Looking at Figure 11, there is a positive correlation between the standard deviation of attitude error $\sigma_q$ and $\delta\hat{\rho}$. Similar to the absolute position, Alg. 1 uses the attitude rotation matrix directly. However, $\sigma_q$ does not have the influence that $\sigma_a$ does. In fact, the noise induced by using keypoints as features nearly overshadows the error introduced by $\sigma_q$, as seen in the right-side plot in Figure 11. 60 arcsecs is $0.0167°$, which results in an along-track error of 0.8773mm for a spacecraft observing Tango from 3.01m, which is the default distance to the target from the $\gamma$ in Table 2. Thus, most pointing sensors available[32] will have little influence on V3DR performance.

### Synchronized vs. Sequential Imaging

For all of the simulations using keypoints, synchronized imaging performs as well as, if not better than, sequential imaging. There is negligible difference between the two when model vertices are used, which indicates that synchronized imaging is correlated with improved keypoint detection and matching. On average, synchronized imaging had 20% lower $\delta\hat{\rho}$ across all simulations using keypoints, or 31% and 10% lower $\delta\hat{\rho}$ for Eros and Tango, respectively.

## CASE STUDIES

### Single-Agent, Non-Cooperative Spacecraft Rendezvous

One of the most common applications of target rendezvous is a single spacecraft approaching a non-cooperative spacecraft whose size, shape, mass, orbit, and rotation rate are unknown. With such

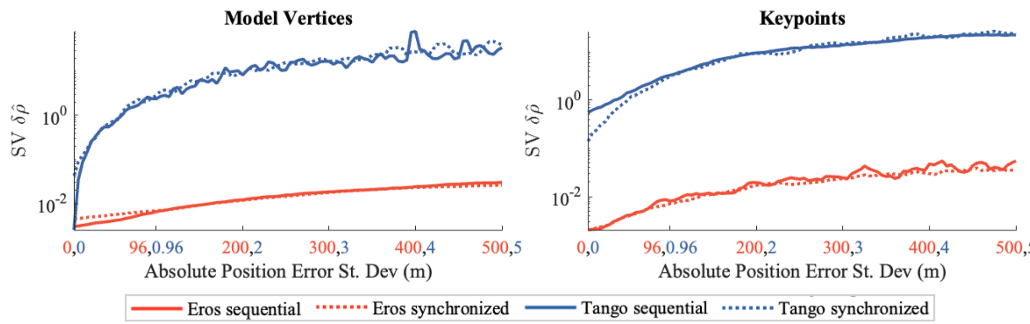



**Figure 8**: The effects of absolute position error $\sigma_a$ on SV performance. The red and blue x-labels correspond to Eros and Tango, respectively.

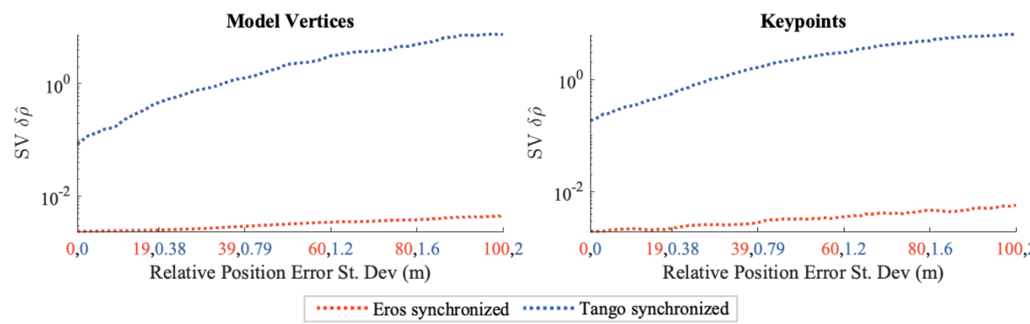



**Figure 9**: The effects of relative position error $\sigma_r$ on SV performance. The red and blue x-labels correspond to Eros and Tango, respectively.

a lack of information, SfM is the recommended to recover the target's shape and the observer's pose. The rotation rate of the target must still be estimated to ensure adequate angular separation between the first and last image, but a rough estimate can be obtained through optical flow techniques.

In this case study, one observer orbits the Tango spacecraft while Tango rotates about a single-axis at $\dot{\theta} = 1°$/s. Three relative geometry parameters must be decided: $\beta$, $N$, and $\gamma$. Based on Figures 4 and 6, the ideal parameter ranges are $\beta \in [20, 40]°$, $N \in [2, 4]$, and $\gamma \in [60, \infty)\%$. SfM begins to fail as $\beta \to 40°$ and $N \to 4$ so $\beta$ and $N$ are chosen to be $20°$ and 3, respectively. The second image also provides indirect matches between the first and last images. SfM performance for *Tango sequential keypoints* plateaus after $\gamma = 60\%$ so ideally $\gamma > 60\%$. However, this results in $d = 3.9$m from Eq. (2), which is very close to the target. Instead, $\gamma$ is chosen to be 40%, which results in $d = 5.9$m.

Figure 12 shows the results of SfM initialization with the proposed setup. For this case, $\delta\hat{\rho} = 0.013383$ and $\delta\hat{P} = 19.807$mm, which agrees which is slightly lower than the $\delta\hat{\rho}$ value for $\gamma = 40\%$ for *Tango sequential keypoints* in Figure 2.

## Multi-Agent Navigation and Characterization of an Asteroid

Asteroid navigation and characterization historically uses high SWaP-C equipment and frequent ground-in-the-loop communication. A multi-agent system that uses V3DR in lieu of LiDAR can increase system redundancy, measurement observability, and autonomy. Telescope observations

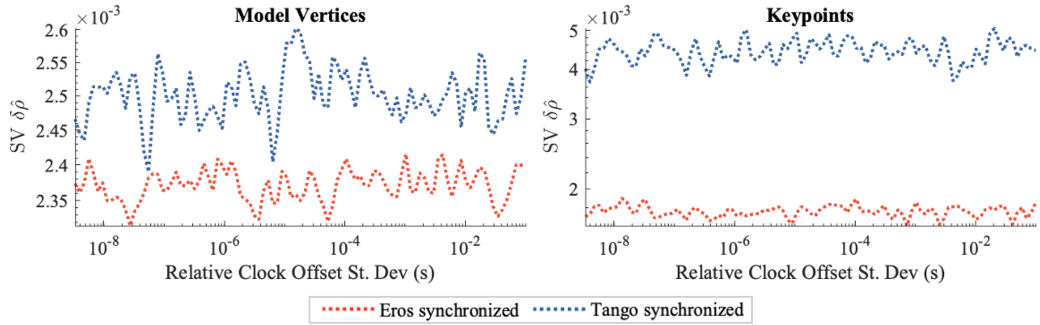



**Figure 10**: The effects of relative clock offset $\delta t_n$ on multi-agent SV performance.

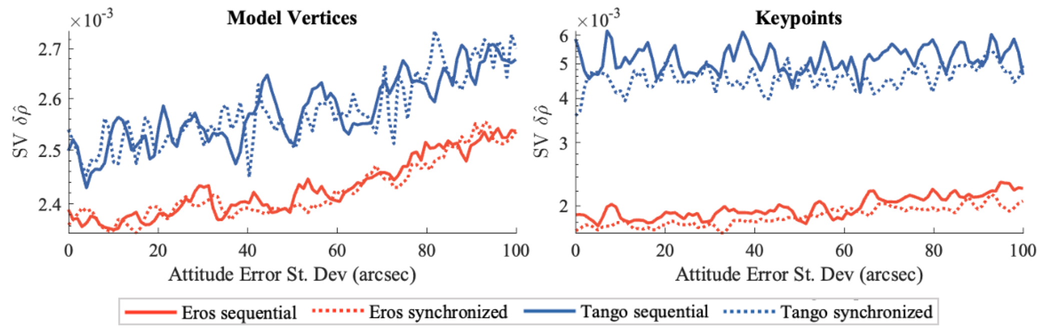



**Figure 11**: The effects of attitude error $\sigma_q$ on SV performance.

and a long approach phase can determine initial estimates for the target and spacecraft states.[3] With such data available, SV is expected to have a lower error than SfM.

In this case study, a DSS orbits asteroid 433 Eros in the same formation described in Table 1 and state uncertainty values listed in Table 3. $N$ is set to 3 and $\beta = 15°$. This $\beta$ increases keypoint matches but using three agents total increases the total angular separation to 30° to maintain V3DR performance. With Eros as the target, $\gamma$ can be increased significantly while maintaining a safe separation from the target so $\gamma = 80\%$, which results in $d = 64.985$km.

Figure 13 shows the results of SV with the proposed setup. For this case, $\delta\hat{\rho} = 3.1452 \times 10^{-3}$ and $\delta\hat{P} = 103.793$m. Like with Tango as the target, this $\delta\hat{\rho}$ is slightly higher than that of $\gamma = 80\%$ in Figure 2 for *Eros synchronized keypoints*. Furthermore, these results agree with a previous publication from the author, Dennison et al. (2021).[17] That publication performs a similar case study for a swarm of three spacecraft orbiting Asteroid 433 Eros using a rigorous dynamics simulation and an unscented Kalman filter for state estimation.

## CONCLUSIONS

Vision-based 3D reconstruction (V3DR) techniques such as stereovision (SV) and structure from motion (SfM) are underutilized in practice in the space domain. They have great potential for use in missions where low SWAP-C hardware is a priority. This paper addressed three knowledge gaps that obscured how V3DR should be implemented in a space mission: 1) V3DR sensitivity to observer state knowledge, number of observers, target motion, and clock synchronization is rarely studied; 2) sequential imaging has not been compared directly to synchronized imaging; and 3) techniques are

**Figure 12**: Results of SfM initialization on three sequential images from a single agent observing a non-cooperative spacecraft. Shown are the true observer poses (green cameras), ray-traced target surface features (green dots), estimated observer poses (red cameras), the reconstructed surface features (red dots), the Tango 3D shape model (gray), and the lighting direction (blue line).

not harmonized between the various space-domain applications. A sensitivity analysis for SV and SfM on various multi-agent and multi-view relative geometry parameters as well as several observer state uncertainty parameters revealed trends in V3DR with respect to these domain gaps.

The results of these simulations determine optimal ranges for image angular separation, the number of views used for V3DR, imaging rate, and the distance to the target. They also show that the order of magnitude improvement of relative position error compared to absolute position error translates directly to SV performance. It is also shown that multi-agent SV is robust to image desynchonization effects of clock synchronization error. Furthermore, V3DR using synchronized images from multiple agents is less sensitive to target rotation rate than V3DR using sequential imaging from a single agent. Synchronized imaging also results in 20% lower error, on average, than sequential imaging.

Finally, two case studies applied the lessons learned from the design parameter simulations. For single-agent SfM performed on a non-cooperative spacecraft, points on the target's surface were reconstructed with a median error of 19.07mm. The second case study performed multi-agent SV with realistic a state noise model and resulted in a median reconstruction error of 103.793m.

## ACKNOWLEDGEMENTS

This work is partially supported by Scientific Systems Company, Inc. (Subcontract No. SC-1761-01), in support of Prime Contract No. FA8649-22-P-1184 with the U.S. Air Force AFWERX's STTR Program, titled:"Unknown Satellite Realtime Inspection Pose, Tracking and IDEntification (US-RIPTIDE)

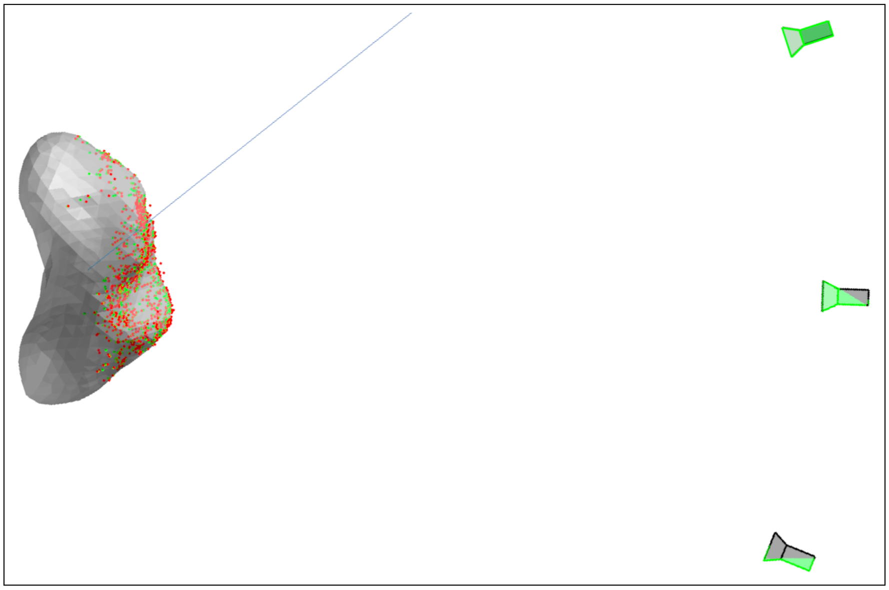


**Figure 13**: Multi-agent, synchronized SV performance of three spacecraft orbiting an asteroid. Shown are the true observer poses (green cameras), ray-traced target surface features (green dots), the noisy observer poses used for SV (black cameras), the reconstructed surface features (red dots), the asteroid 3D shape model (gray), and the lighting direction (blue line).

## REFERENCES


[1] K. Dennison, N. Stacey, and S. D'Amico, "Autonomous Asteroid Characterization Through Nanosatellite Swarming," *IEEE Transactions on Aerospace and Electronic Systems*, 2023, pp. 1–22, 10.1109/TAES.2023.3245997.

[2] D. A. Lorenz, R. Olds, A. May, C. Mario, M. E. Perry, E. E. Palmer, *et al.*, "Lessons learned from OSIRIS-REx autonomous navigation using natural feature tracking," *IEEE Aerospace Conference*, Big Sky, MT, USA, Mar. 2017, 10.1109/AERO.2017.7943684.

[3] B. G. Williams, "Technical challenges and results for navigation of NEAR Shoemaker," *Johns Hopkins APL Technical Digest (Applied Physics Laboratory)*, 2002.

[4] T. H. Park and S. D'Amico, "Adaptive Neural Network-based Unscented Kalman Filter for Spacecraft Pose Tracking at Rendezvous," 2022, 10.48550/ARXIV.2206.03796.

[5] P. Rousso, S. Samsam, and R. Chhabra, "A Mission Architecture for On-Orbit Servicing Industrialization," *2021 IEEE Aerospace Conference*, IEEE, 2021, pp. 1–14.

[6] R. Opromolla, G. Fasano, G. Rufino, and M. Grassi, "A review of cooperative and uncooperative spacecraft pose determination techniques for close-proximity operations," *Progress in Aerospace Sciences*, Vol. 93, 2017, pp. 53–72.

[7] R. W. Gaskell, O. S. Barnouin-Jha, D. J. Scheeres, A. S. Konopliv, T. Mukai, S. Abe, *et al.*, "Characterizing and navigating small bodies with imaging data," *Meteoritics & Planetary Science*, Vol. 43, June 2008, pp. 1049–1061, 10.1111/j.1945-5100.2008.tb00692.x.

[8] S. Watanabe, M. Hirabayashi, N. Hirata, N. Hirata, R. Noguchi, Y. Shimaki, *et al.*, "Hayabusa2 arrives at the carbonaceous asteroid 162173 Ryugu—A spinning top–shaped rubble pile," *Science*, Vol. 364, Apr. 2019, pp. 268–272, 10.1126/science.aav8032.

[9] S. Sharma and S. D'Amico, "Neural network-based pose estimation for noncooperative spacecraft rendezvous," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 56, No. 6, 2020, pp. 4638–4658.

[10] M. Driedger, A. Asgari, and P. Ferguson, "Feasibility of Gathering Resident Space Object Range Measurements Using In-Orbit Observers," *IEEE Journal of Radio Frequency Identification*, 2022.

[11] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge: Cambridge University Press, 2003.

[12] S. Nousias, M. Lourakis, and C. Bergeles, “Large-scale, metric structure from motion for unordered light fields,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3292–3301.

[13] S. Henry and J. A. Christian, “Absolute Triangulation Algorithms for Space Exploration,” *Journal of Guidance, Control, and Dynamics*, Vol. 46, No. 1, 2023, pp. 21–46.

[14] J. Kruger, K. Wallace, A. Koenig, and S. D’Amico, “Autonomous Angles-Only Navigation for Spacecraft Swarms around Planetary Bodies,” *IEEE Aerospace Conference*, Virtual, 2021.

[15] M. Dor and P. Tsiotras, “ORB-SLAM applied to spacecraft non-cooperative rendezvous,” *AAS/AIAA Space Flight Mechanics Meeting*, 2018.

[16] M. R. U. Saputra, A. Markham, and N. Trigoni, “Visual SLAM and Structure from Motion in Dynamic Environments: A Survey,” *ACM Computing Surveys*, Vol. 51, June 2018, pp. 1–36, 10.1145/3177853.

[17] K. Dennison and S. D’Amico, “Comparing Optical Tracking Techniques in Distributed Asteroid Orbiter Missions Using Ray-Tracing,” *AAS/AIAA Space Flight Mechanics Meeting*, Charlotte, NC, Feb. 2021.

[18] N. Takeishi, A. Tanimoto, T. Yairi, Y. Tsuda, F. Terui, N. Ogawa, and Y. Mimasu, “Evaluation of Interest-region Detectors and Descriptors for Automatic Landmark Tracking on Asteroids,” *Transactions of the Japan Society for Aeronautical and Space Sciences*, Vol. 58, No. 1, 2015, pp. 45–53, 10.2322/tjsass.58.45.

[19] M. Hess-Flores, S. Recker, and K. I. Joy, “Uncertainty, Baseline, and Noise Analysis for L1 Error-Based Multi-view Triangulation,” *22nd International Conference on Pattern Recognition*, Stockholm, Sweden, Aug. 2014, pp. 4074–4079, 10.1109/ICPR.2014.698.

[20] O. Enqvist, F. Kahl, and C. Olsson, “Non-sequential structure from motion,” *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, IEEE, 2011, pp. 264–271.

[21] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, “ORB-SLAM: a versatile and accurate monocular SLAM system,” *IEEE transactions on robotics*, Vol. 31, No. 5, 2015, pp. 1147–1163.

[22] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, Vol. 60, Nov. 2004, pp. 91–110, 10.1023/B:VISI.0000029664.99615.94.

[23] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” *2011 International Conference on Computer Vision*, Barcelona, Spain, Nov. 2011, pp. 2564–2571, 10.1109/ICCV.2011.6126544.

[24] L. P. Cassinis, R. Fonod, E. Gill, I. Ahrns, and J. G. Fernandez, “Comparative assessment of image processing algorithms for the pose estimation of uncooperative spacecraft,” *Proc. Int. Workshop Satell. Constellations Formation Flying*, 2019, pp. 1–20.

[25] D. Nistér, “An efficient solution to the five-point relative pose problem,” *IEEE transactions on pattern analysis and machine intelligence*, Vol. 26, No. 6, 2004, pp. 756–770.

[26] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, “Complete solution classification for the perspective-three-point problem,” *IEEE transactions on pattern analysis and machine intelligence*, Vol. 25, No. 8, 2003, pp. 930–943.

[27] M. I. Lourakis and A. A. Argyros, “SBA: A software package for generic sparse bundle adjustment,” *ACM Transactions on Mathematical Software (TOMS)*, Vol. 36, No. 1, 2009, pp. 1–30.

[28] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, “Generic and real-time structure from motion using local bundle adjustment,” *Image and Vision Computing*, Vol. 27, No. 8, 2009, pp. 1178–1193.

[29] M. D’Errico, *Distributed space missions for earth system monitoring*, Vol. 31. Springer Science & Business Media, 2012.

[30] E. Hecht, *Optics, global edition*. Pearson Education UK, 2016.

[31] O. Montenbruck, E. Gill, and F. Lutze, “Satellite orbits: models, methods, and applications,” *Appl. Mech. Rev.*, Vol. 55, No. 2, 2002, pp. B27–B28.

[32] C. Pong, “On-Orbit Performance & Operation of the Attitude & Pointing Control Subsystems on ASTERIA,” *32nd Annual Small Satellite Conference*, Utah, USA, Aug. 2018.

[33] S. D’Amico, *Autonomous formation flying in low earth orbit*. PhD thesis, TU Delft, 2010.

[34] R. C. Gonzalez and R. E. Woods, *Digital image processing*. Pearson, 2018.

[35] P. Patidar, M. Gupta, and A. K. Nagawat, “Image Denoising by Various Filters for Different Noise Image De-noising by Various Filters for Different Noise Sumit Srivastava,” 2010.

[36] FLIR, “Grasshopper3 USB3 Teledyne FLIR,”