Demystifying Multi-Agent Stereovision in the Asteroid Rendezvous Environment

Kaitlin Dennison
Stanford University
Stanford, California 94305
kdenn@stanford.edu

Abstract

Noise and error tolerances of multi-agent stereovision methods are assessed in the context of an asteroid rendezvous mission. Enabling low-power, low-computation, and readily-available alternatives to LiDAR and stereophotoclinometry are critical to enable future missions to asteroids. Triangulation and outlier rejection methods are assessed with varying sensor noise and camera relative position parameters. Simulations were performed using OpenGL generated images of Asteroid 433 Eros and modeling two identical cameras observing the target. SIFT keypoints and triangular mesh face centers were used as landmarks. Nonlinear triangulation with point pre-conditioning performed the best of the triangulation methods but only showed marginal improvement over linear triangulation with point pre-conditioning. The epipolar constraint and MLESAC with the fundamental matrix and the essential matrix as models all performed comparatively as outlier rejection methods. All were highly sensitive to parameter tuning. There is an inverse relationship between increasing angular separation between cameras and triangulation depth error. This competes with the nature of feature correlation to perform better with smaller angular separation. In terms of noise, camera position error has the greatest influence on triangulation depth error. The effects of pixel measurement and star tracker uncertainty were minimal compared to those from camera position error.

1. Introduction

Understanding asteroids has the potential to advance research in off-Earth resource gathering, asteroid-Earth collision prevention, and many other areas of scientific study [1, 23]. However, almost all asteroid rendezvous missions to date use large, monolithic spacecraft with custom hardware built for the mission. They also heavily rely on downlinking data to Earth for processing because of the limited computation power onboard spacecraft [18, 20, 28].

A possible solution is to use fully commercial off-the-shelf hardware, such as CubeSatellite buses, star trackers [21], and monocular cameras [8] to perform multi-agent stereovision with a swarm of nanosatellites. This could eliminate the need for power-intensive LiDAR or waiting for stereophotoclinometry from Earth [7]. The purpose of this paper is to evaluate the error and noise tolerances of various stereovision algorithms to determine the requirements of such a system.

In fact, a recently published study [25] showed autonomous simultaneous navigation and characterization of asteroid Eros 433 using multi-agent stereovision to potentially be feasible. One sticking point, however, was that stereovision estimation could not use the mean state estimate from the navigation filter or the filter estimates would diverge. Instead, the state provided to stereovision was the true state plus Gaussian noise drawn from the filter covariance. Thus, this paper also aims to shed light on why this occurred.

The structure of the remaining paper is as follows. Section 2 details mathematical algorithms for sub-components necessary to perform stereovision. Section 3 explains the experimental setup, which draws inspiration from [25]. Section 4 discusses the results of the simulations. Finally, conclusions are drawn in Section 5 based on the results.

2. Mathematical Preliminaries

The camera projection matrix to the image frame must be defined. The world frame used is the asteroid-centered asteroid-fixed (ACAF) frame, which is defined by the International Astronomical Union for each specific asteroid [2]. The origin of the ACAF frame is the asteroid center of mass and the frame rotates with the asteroid. A landmark 3D position, $\boldsymbol{X}$, is projected from the ACAF frame into the image frame of the jth spacecraft using the finite projective camera model [9],

$$\begin{bmatrix} uw \\ vw \\ w \end{bmatrix} = \mathbf{K}_j \begin{bmatrix} \underset{ACAF \to CF_j}{\mathbf{R}} & \underset{ACAF \to CF_j}{\mathbf{R}}(-\boldsymbol{r}_j) \end{bmatrix} \begin{bmatrix} \boldsymbol{X} \\ 1 \end{bmatrix}. \tag{1}$$

The projected pixel measurements of the landmark center,

$$\boldsymbol{x}_j = \begin{bmatrix} u \\ v \end{bmatrix}, \tag{2}$$

are computed by dividing the first two elements of Eq. (1) by w, which is the common factor of the homogeneous coordinate of $\boldsymbol{x}$. The rotation matrix from the ACAF frame to the camera-fixed (CF) frame of the jth spacecraft is ${}_{ACAF\rightarrow CF_j}\mathbf{R}$, which is provided by the star tracker onboard that spacecraft. The ACAF vector of the jth spacecraft's center of mass relative to the asteroid center of mass is $\boldsymbol{r}_j$. Finally, the camera intrinsic matrix is,

$$\mathbf{K}_j = \begin{bmatrix} k_x & \gamma & c_x \\ 0 & k_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

where k_x and k_y are the camera focal length divided by the pixel pitch in the CF frame x and y directions, respectively. The principal point in pixels is denoted $\boldsymbol{c} = [c_x,\ c_y]^T$. The scalar, γ, is a skew factor generally determined from camera calibration.

Equation (1) can be reduced to a concise form,

$$\boldsymbol{x}_{j,h} = \mathbf{M}_j \boldsymbol{X}_h. \tag{4}$$

The h subscript denotes the homogeneous form of a point and

$$\mathbf{M}_j = \mathbf{K}\begin{bmatrix} \underset{ACAF\rightarrow CF_j}{\mathbf{R}} & \underset{ACAF\rightarrow CF_j}{\mathbf{R}}(-\boldsymbol{r}_j) \end{bmatrix} \tag{5}$$

is the camera projective transform of the jth spacecraft.

2.1. Epipolar Geometry

The relationship between two camera views of the same landmark is defined by their epipolar geometry. The epipolar plane is the 3D plane connecting the two camera optical centers and $\boldsymbol{X}$. This plane intersects the image frame of both cameras, forming an epipolar line, $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$, in each image that intersects the respective 2D projection of $\boldsymbol{X}$: $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ [9].

2.1.1 The Fundamental Matrix

A 3×3, rank 2 matrix, the fundamental matrix $\mathbf{F}$, relates the two image frames. This matrix has seven degrees of freedom and is only defined up to a scale [9]. It is related to the camera matrices via,

$$\mathbf{F} = [\boldsymbol{e}_2]_\times \mathbf{M}_2 \mathbf{M}_1^\dagger \tag{6}$$

where $\boldsymbol{e}_2$ is the epipole in the second image, which is the homogeneous projection of the first camera's optical center in the second camera's image frame. Furthermore, $[\boldsymbol{e}_2]_\times$ is the skew-symmetric form of $\boldsymbol{e}_2$ and $\mathbf{M}_1^\dagger$ is the pseudo-inverse of $\mathbf{M}_1$.

$\mathbf{F}$ relates the image points to each other and the image points to the epipolar lines through

$$\boldsymbol{x}_{2,h}^T \mathbf{F} \boldsymbol{x}_{1,h} = 0, \tag{7}$$

$$\boldsymbol{l}_2 = \mathbf{F}\boldsymbol{x}_{1,h}, \tag{8}$$

and

$$\boldsymbol{l}_1 = \mathbf{F}^T \boldsymbol{x}_{2,h}. \tag{9}$$

It is possible to estimate $\mathbf{F}$ without knowledge of the camera matrices or epipole. One method is the *eight-point algorithm*. Because $\mathbf{F}$ has seven degrees of freedom plus a scale ambiguity, Eq. (7) can be leveraged with eight corresponding image pairs to provide the eight constraints needed to solve for $\mathbf{F}$. These constraints are reconfigured into a linear form,

$$\mathbf{W}\boldsymbol{f} = \mathbf{0}. \tag{10}$$

Here, $\boldsymbol{f}$ is the column-form of $\mathbf{F}$ and the ith row of $\mathbf{W}$ is the vector, $[u_1u_2, v_1u_2, u_2, u_1v_2, v_1v_2, v_2, u_1, v_1, 1]$. The ith row of $\mathbf{W}$ uses the ith pair of corresponding image points so there are n rows for n pairs of points.

The vector $\boldsymbol{f}$ is solved for via singular value decomposition (SVD) of $\mathbf{W}$. The pixel measurements provided are often noisy, leading to a full rank estimate for $\mathbf{F}$. In this case, SVD is employed again to set the third singular value to zero and re-composing $\mathbf{F}$ [9]. An important modification to the eight-point algorithm will be discussed in Section 2.3 Point Pre-Conditioning.

2.1.2 The Essential Matrix

A similar matrix exists for the calibrated camera case: the essential matrix $\mathbf{E}$. This matrix is also 3×3, rank 2, and of ambiguous scale, but has only five degrees of freedom. The fundamental and essential matrices are related by the two camera intrinsic matrices,

$$\mathbf{E} = \mathbf{K}_2^T \mathbf{F} \mathbf{K}_1. \tag{11}$$

$\mathbf{E}$ provides similar relationships between points and epipolar lines as $\mathbf{F}$, but $\mathbf{E}$ operates under the assumption of *normalized coordinates* where $\hat{\boldsymbol{x}}_{j,h} = \mathbf{K}_j \boldsymbol{x}_{j,h}$. The point and epipolar line relationships are,

$$\hat{\boldsymbol{x}}_{2,h}^T \mathbf{E} \hat{\boldsymbol{x}}_{1,h} = 0, \tag{12}$$

$$\hat{\boldsymbol{l}}_2 = \mathbf{E}\hat{\boldsymbol{x}}_{1,h}, \tag{13}$$

and

$$\hat{\boldsymbol{l}}_1 = \mathbf{E}^T \hat{\boldsymbol{x}}_{2,h}. \tag{14}$$

The matrix $\mathbf{E}$ has two constraints characterized by,

$$2\mathbf{E}\mathbf{E}^T\mathbf{E} - tr(\mathbf{E}\mathbf{E}^T)\mathbf{E} = 0. \tag{15}$$

Furthermore, because $\mathbf{E}$ and $\mathbf{E}$ are rank 2, $\det(\mathbf{E}) = \det(\mathbf{F}) = 0$. These constraints combined with Eq. (7) allow just five points to be used to estimate $\mathbf{E}$. Derivations of modern solutions to this problem are in [13] and [15].

2.1.3 Outlier Rejection

When performing computer vision involving corresponding pairs of points between multiple views, there will often be spurious matches. These outliers must be identified and removed from the dataset to obtain the best possible results of stereovision, structure from motion, or many other computer vision applications. While there are countless methods to choose from for outlier rejection, two are explained here that take advantage of the relationships introduced in the previous parts of Section 2.1 Epipolar Geometry.

The first method is the epipolar constraint. If two image points correspond to the same $\boldsymbol{X}$, the epipolar line $\boldsymbol{l}_2$ in the second image projected from the point $\boldsymbol{x}_1$ (see Eq. (8)) in the first image will intersect $\boldsymbol{x}_2$ in the second image. The reverse is also true for the other epipolar line $\boldsymbol{l}_1$. In the presence of noise, $\boldsymbol{l}_2$ and $\boldsymbol{x}_2$ may not intersect, but $\boldsymbol{x}_2$ should not be too far away. If the line is parameterized as $\boldsymbol{l}_2 = [a, b, c]^T$ such that $au + bv + c = 0$ for a point (u, v) on the line, the distance d from $\boldsymbol{x}_2$ to $\boldsymbol{l}_2$ is, [16]

$$d = \frac{|\boldsymbol{l}_2^T \boldsymbol{x}_{2,h}|}{\sqrt{a^2 + b^2}}. \tag{16}$$

Thus, a distance threshold can be set to identify outlier pairs of image points. The fundamental or essential matrix can be used to determine the epipolar lines.

The second method is random sample consensus (RANSAC), which randomly samples a subset of provided data points to estimate a model that could have resulted in those data points [6]. There are three steps to RANSAC that are iterated over until some termination criterion is met. One, select a random sample of data points, at least as many as are needed to fit the model. Two, compute a model from that subset. Determine which points from the entire subset are inliers to that model. Typical stopping criterion are percentage of inliers over the entire dataset, the model fit residual, and simply iteration count. There has been a more recent development to move to MLESAC, which uses the maximum likelihood estimate in lieu of the number of inliers [26].

There are two models that relate image point pairs that have already been discussed in this paper: the fundamental and essential matrices. There is a third possible model under the condition that either all 3D points lie on the same plane or the camera rotates but does not translate: a homography such that $\boldsymbol{x}_{2,h} = \mathbf{H}\boldsymbol{x}_{1,h}$. The MLESAC formulations for all three models are discussed in [26].

2.2. Triangulation

Structure from motion is the estimation of image depth information (structure) in addition to the camera view parameters (motion). By contrast, *stereovision* is the general process of recovering the depth from image points from two known camera views. The method of computing a 3D point estimate from correlated image points from any number of views is called *triangulation*. There are multiple ways to perform triangulation [10, 17], but most stem from linear triangulation. Thus, linear triangulation is described here in the context of this paper.

Given the matched pixel measurements from two spacecraft and their respective projective transforms, two equations can be formed from Eq. (4),

$$\boldsymbol{x}_{1,h} = \mathbf{M}_1 \boldsymbol{X}_h \tag{17}$$

$$\boldsymbol{x}_{2,h} = \mathbf{M}_2 \boldsymbol{X}_h. \tag{18}$$

Because the cross product of two parallel vectors is zero, two more equations can be formed,

$$\boldsymbol{x}_{1,h} \times \mathbf{M}_1 \boldsymbol{X}_h = \mathbf{0} \tag{19}$$

$$\boldsymbol{x}_{2,h} \times \mathbf{M}_2 \boldsymbol{X}_h = \mathbf{0}. \tag{20}$$

Equations (17 - 20) share the same 3D landmark position and can be combined into a form, $\mathbf{A}\boldsymbol{X}_h = \mathbf{0}$, where

$$\mathbf{A} = \begin{bmatrix} u_1 \boldsymbol{M}_1^{(3)T} - \boldsymbol{M}_1^{(1)T} \\ v_1 \boldsymbol{M}_1^{(3)T} - \boldsymbol{M}_1^{(2)T} \\ u_2 \boldsymbol{M}_2^{(3)T} - \boldsymbol{M}_2^{(1)T} \\ v_2 \boldsymbol{M}_2^{(3)T} - \boldsymbol{M}_2^{(2)T} \end{bmatrix}. \tag{21}$$

Each $\boldsymbol{M}_j^{(r)T}$ is the is the rth row of $\mathbf{M}_j$. There is likely no exact solution to $\mathbf{A}\boldsymbol{X}_h = 0$ because there is noise in both $\boldsymbol{x}_{j,h}$ and $\mathbf{M}_j$. Instead, $\boldsymbol{X}_h$ is considered the minimizer of $\|\mathbf{A}\boldsymbol{X}_h\|_2$ subject to $\|\boldsymbol{X}_h\|_2 = 1$. The constraint is included because $\boldsymbol{X}_h = \mathbf{0}$ is a trivial solution. The matrix $\mathbf{M}_j$ is only defined up to a scale, so the value of $\|\boldsymbol{X}_h\|_2$ is unimportant. This optimization problem is solved via singular value decomposition where $\boldsymbol{X}_h$ is the unit right-singular vector corresponding to the smallest singular value of $\mathbf{A}$ [9].

Linear triangulation can be extended to $N > 2$ views by appending additional rows onto $\mathbf{A}$. One downside to this formulation, however, is that each 3D point must be processed individually. Another caveat is that it is not projective invariant [10].

2.2.1 Nonlinear Triangulation

Linear triangulation estimates can be improved by solving the system as a minimization problem over the reprojection error. This problem is formulated as,

$$\min_{\tilde{\boldsymbol{X}}} \|\mathbf{M}_1\tilde{\boldsymbol{X}}_h - \boldsymbol{x}_{1,h}\| + \|\mathbf{M}_2\tilde{\boldsymbol{X}}_h - \boldsymbol{x}_{2,h}\|. \quad (22)$$

The variable $\tilde{\boldsymbol{X}}_h$ is the estimated 3D point in homogeneous coordinates. More cameras can be included by simply adding the reprojection error on to Eq. (22). The more general form is,

$$\min_{\tilde{\boldsymbol{X}}} \sum_{j=1}^{N} \left(\mathbf{M}_j\tilde{\boldsymbol{X}} - \boldsymbol{x}_j\right). \quad (23)$$

This is formulated as a nonlinear least squares problem and can be solved using optimization techniques. An initial guess for $\tilde{\boldsymbol{X}}_h$ is required and is typically obtained via linear triangulation.

A common technique is to use the Gauss Newton method, which requires the Jacobian $\mathbf{J}$ of the cost function in 23 [9].

$$\mathbf{J} = -\begin{bmatrix} \frac{\mathbf{M}_j^{(31)}\mathbf{M}_j^{(1)}\tilde{\boldsymbol{X}}_h}{(\mathbf{M}_j^{(3)}\tilde{\boldsymbol{X}}_h)^2} & \frac{\mathbf{M}_j^{(32)}\mathbf{M}_j^{(1)}\tilde{\boldsymbol{X}}_h}{(\mathbf{M}_j^{(3)}\tilde{\boldsymbol{X}}_h)^2} & \frac{\mathbf{M}_j^{(33)}\mathbf{M}_j^{(1)}\tilde{\boldsymbol{X}}_h}{(\mathbf{M}_j^{(3)}\tilde{\boldsymbol{X}}_h)^2} \\ \frac{\mathbf{M}_j^{(31)}\mathbf{M}_j^{(2)}\tilde{\boldsymbol{X}}_h}{(\mathbf{M}_j^{(3)}\tilde{\boldsymbol{X}}_h)^2} & \frac{\mathbf{M}_j^{(32)}\mathbf{M}_j^{(2)}\tilde{\boldsymbol{X}}_h}{(\mathbf{M}_j^{(3)}\tilde{\boldsymbol{X}}_h)^2} & \frac{\mathbf{M}_j^{(33)}\mathbf{M}_j^{(2)}\tilde{\boldsymbol{X}}_h}{(\mathbf{M}_j^{(3)}\tilde{\boldsymbol{X}}_h)^2} \end{bmatrix} + \frac{1}{\mathbf{M}_j^{(3)}\tilde{\boldsymbol{X}}_h}\begin{bmatrix}\mathbf{M}_j^{(1)} \\ \mathbf{M}_j^{(2)}\end{bmatrix}. \quad (24)$$

$\mathbf{M}_j^{(rc)}$ is the cth element of the rth row of $\mathbf{M}_j$ and $\mathbf{M}_j^{(r)}$ is the entire rth row of $\mathbf{M}_j$.

Other methods of solving nonlinear triangulation can be found in [9, 10, 17].

2.3. Point Pre-Conditioning

There is a simple transform that can be applied to the 2D image points that aids almost every estimation algorithm discussed up to this point. In a matrix like $\mathbf{W}$ in Eq. (10), the scale of the individual elements can vary widely because of the multiplications and the necessary 1's in the last column. This leads to $\mathbf{W}$ being ill-conditioned for SVD, which works best if $\mathbf{W}$ has one singular value close (or equal) to zero and the others nonzero.

The solution is to pre-condition the points in an image by applying a translation and scaling. First, the points are translated such that the origin of the new coordinate system is at the centroid of the image point. Second, the points are scaled so the mean square distance of the translated points is 2 pixels. The transform can be applied with a single matrix,

$$\mathbf{T} = \begin{bmatrix} s & 0 & -s\bar{u} \\ 0 & s & -s\bar{v} \\ 0 & 0 & 1 \end{bmatrix}, \quad (25)$$

where,

$$s = \left(\frac{2n}{\sum_{i=1}^{n}\|\boldsymbol{x}^{(i)} - \bar{\boldsymbol{x}}\|^2}\right)^{1/2}. \quad (26)$$

Here, n is the number of points in the image, $\boldsymbol{x}^{(i)}$ is the ith image point, $\bar{\boldsymbol{x}}$ is the mean of the image points, and $\bar{u}$ and $\bar{v}$ are the mean of the x- and y-components of all the image points, respectively [9]. The pre-conditioned points are,

$$\breve{\boldsymbol{x}}_{j,h} = \mathbf{T}_j\boldsymbol{x}_{j,h}. \quad (27)$$

This process can return more accurate estimates from the eight-point algorithm and is called the *normalized* eight-point algorithm. It also improves results from essential matrix estimation and linear triangulation so it is generally recommended to pre-condition your points when generating estimates from large sets of image points [9]. A matrix computed using pre-conditioned points, like the fundamental matrix $\breve{\mathbf{F}}$, can be de-normalized by

$$\mathbf{F} = \mathbf{T}_2\breve{\mathbf{F}}\mathbf{T}_1. \quad (28)$$

Pre-conditioning points is often referred to as *normalizing* the points, but note that it is not the same as the normalized coordinates discussed in Section 2.1.2 The Essential Matrix.

3. Simulation Design

The main focus of these simulations is to discern the relationship between three aspects of a mission to an asteroid. One, the relative position and orientation of two spacecraft as well as of the spacecraft and the asteroid. Two, algorithms used to aid or perform stereovision. Three, noise and error sources from sensors or state estimation. Parameters for each aspect of position, noise, and error are summarized in Table 1.

All simulations will use asteroid 433 Eros as the target because of the great amount of high-fidelity data available from the NEAR Shoemaker mission [14, 31]. Eros has a mean radius of 8.42 km [30]; some results will be presented in terms of distance units (1 DU = 8.42 km) so these findings can more easily be applied to other asteroids or targets of varying sizes.

Images are generated in MATLAB using OpenGL, the ephemeris of Eros [22], the Eros 3D model from the NEAR Shoemaker mission[14], and the camera model in Eq. (1). Simulated images incorporate incidence angle shading and soft shadows but no noise or skew [11, 24]. All spacecraft are modeled with the same short-range camera: the GOMSpace NanoCam with an 8 mm lens [8].

3.1. Relative Position

Figure 1 shows the experimental setup with two cameras. The cameras are primarily separated in the CF_1 X-direction

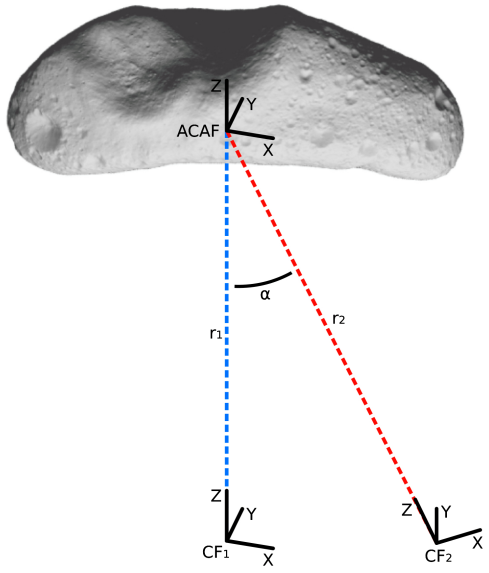

Figure 1. Simulation setup with two cameras. Both cameras are the same distance from Eros' center of mass but camera 2 is rotated -α about the ACAF Y-axis. CF_1 is aligned with ACAF but translated in the ACAF -Z-direction.

with a marginal separation in the Z-direction. CF_2 is rotated about the optical center to maintain a direct pointing at the asteroid center of mass. Thus, the rotation matrices are,

$$\underset{ACAF \to CF_1}{\mathbf{R}} = \mathbf{I}_{3\times 3} \tag{29}$$

and

$$\underset{ACAF \to CF_2}{\mathbf{R}} = \mathbf{R}_Y(\alpha). \tag{30}$$

$\mathbf{R}_Y(\alpha)$ is a passive rotation matrix of α about the y-axis. The ACAF position vectors of the two cameras are,

$$\boldsymbol{r}_1 = \begin{bmatrix} 0 \\ 0 \\ -a \end{bmatrix} \tag{31}$$

and

$$\boldsymbol{r}_2 = \mathbf{R}_Y(-\alpha) \begin{bmatrix} 0 \\ 0 \\ -a \end{bmatrix} \tag{32}$$

where a is the distance of the spacecraft from the asteroid center of mass.

The distance to the target surface influences the magnification of features in the image as well as the percent of the frame that the asteroid occupies. As the spacecraft moves farther from the surface, unique landmarks become closer together in the image plane, making it more difficult to discern one from another.

Similarly, increasing the angle between the cameras increases the perspective change between the images and the baseline distance. A large perspective change can diminish the correlation between image points. Conversely, if the baseline becomes too small, error in the triangulation 3D estimates increases significantly [3, 12].

Thus, the simulations seek to identify how varying a and α influences the results of stereovision. The minimum a is set to 2.13 DU as the maximum radius of Eros is actually 2.10 DU (17.65 km). The maximum a is 8.91 DU, which is where the details on Eros' surface become difficult to observe with the camera model used. Additionally, α is set between 3^o and 30^o. Many of the matrices used in SVD computations become nearly singular for $\alpha < 3^o$ and $\alpha > 30^o$ produced diminishing returns.

3.2. Algorithmic Considerations

There are four algorithmic considerations for performing stereovision being evaluated in these simulations. Three have already been introduced in Section 2 Mathematical Preliminaries: outlier rejection, triangulation, and preconditioning. The fourth is the choice of landmark definition.

The outlier rejection methods that are tested are the epipolar constraint and MLESAC with the fundamental matrix and essential matrix as models. The epipolar constraint was implemented using functions provided by [4], which are MATLAB implementations of various algorithms in [9]. The MLESAC functions are from MATLAB's Computer Vision Toolbox. Furthermore, linear and nonlinear triangulation will be evaluated for point depth reconstruction. These functions are also from [4]. All outlier rejection and triangulation methods are tested with and without point pre-conditioning as detailed in Section 2.3 Point Pre-Conditioning. The only exception is MLESAC with the essential matrix formulation as the MATLAB implementation does not allow for pre-conditioning.

The two landmark definitions for $\boldsymbol{x}$ used are the face centers of a 1708-face triangular mesh model of Eros [14] that are visible to both spacecraft and scale-invariant feature transform (SIFT) keypoint centers [19]. Using the face centers allows the 3D positions and image point correlations to be known exactly. A comparison of keypoint descriptors for asteroid optical navigation can be found in [5]. That paper showed SIFT keypoints to be the most consistent in the asteroid environment. There is no absolute truth ACAF position associated with any keypoint descriptor so the ray-traced 3D points are used as the SIFT ground truth as in [5]. SIFT is implemented using the VLFeat MATLAB Toolbox [27]. Examples of the landmarks are displayed in Figure 2; Figure 3 shows the results of SIFT correlation.

3.3. Noise and Error Sources

A study on asteroid characterization using SIFT [25] showed SIFT features to have approximately 2 px standard

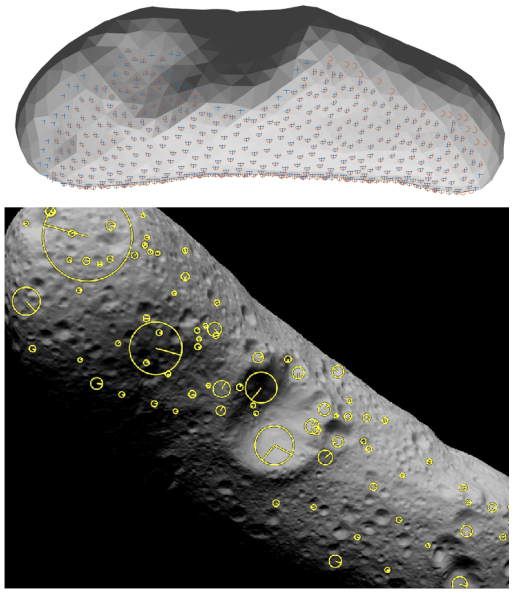

Figure 2. (Top) Eros face centers visible to the first camera (blue plus signs) and second camera (red circle) from $a = 45$ km and $\alpha = 15^{\circ}$. (Bottom) An example OpenGL image of Eros with SIFT keypoints from $a = 35$ km.

deviation of noise for a 45 km orbit about Eros. Thus, noise added to the Eros face centers will range from 0 to 20 px. This noise is scaled with $a/45km$ to account for magnification. When using SIFT features, however, the noise comes inherently from the features themselves so no artificial noise is added to them. Instead, α is varied because SIFT matching reliability decreases as α increases and, consequently, the perspective changes.

The spacecraft position is often estimated using the Deep Space Network's Doppler and range measurements. The NEAR Shoemaker mission to Eros reported errors in their spacecraft position from these measurements to be 100 m [29]. Thus, noise in $\boldsymbol{r}_j$ will be assessed up to 200 m standard deviation.

Error in the star tracker attitude solutions are simulated by multiplying each truth ACAF to CF rotation matrix by a stochastic 3-1-2 Euler angle rotation sequence. The Blue Canyon Technologies Nano-Star Camera [21] is used as a base noise model with the angle of each rotation being a Gaussian random variable with a standard deviation of 24 arcsec for the z-axis rotation and 7 arcsec for the x-axis and y-axis rotations. Simulations use a multiplier m_σ on these base noise levels to maintain consistent scaling:

$$\Sigma_R = \begin{bmatrix} (7m_\sigma)^2 & 0 & 0 \\ 0 & (7m_\sigma)^2 & 0 \\ 0 & 0 & (24m_\sigma)^2 \end{bmatrix} \tag{33}$$

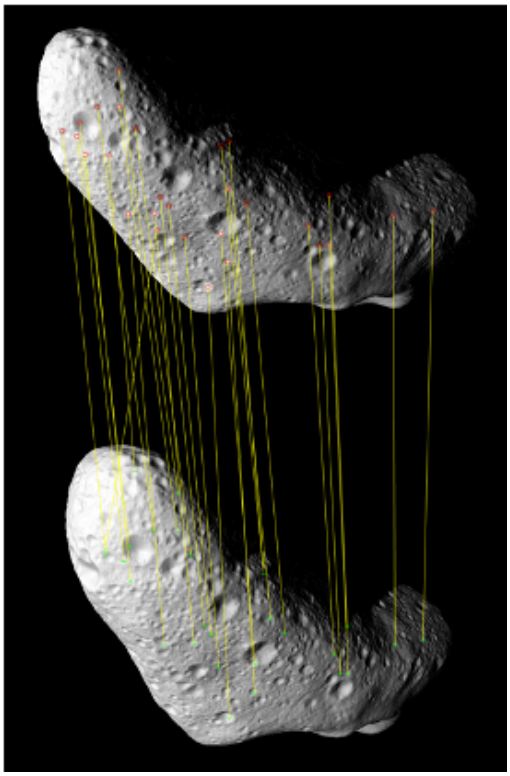

Figure 3. Two OpenGL images of Eros from $a = 45$ km, $\alpha = 10^{\circ}$, with SIFT keypoints matched between them.

Table 1. Noise and Relative Position Parameters

Source	Minimum	Nominal	Maximum
α [$^{\circ}$]	3	15	30
a [DU (km)]	2.13 (18)	5.34 (45)	8.91 (75)
$\boldsymbol{x}$ σ [px]	0	2	20
$\boldsymbol{r}_j$ σ [m]	0	100	200
$\mathbf{R}_{ACAF \to CF}$ m_σ	0	1	20

4. Results and Discussion

Many of the results figures are displayed after the references where there is ample room for comparison.

4.1. Outlier Rejection

Outlier rejection was evaluated by adding artificial miscorrelations to Eros face center pairs or by simply increasing the camera separation α for SIFT keypoints. Figures 6-9 show the results for both types of landmarks as well as with and without nominal noise applied to the system. The solid line is the proportion of true outliers that were rejected. The dash-dot line is the proportion of rejected pairs that were actually outliers. Ideally, both metrics are close to one. PC denotes pre-conditioning.

Adding noise to the system results in significant changes for the face centers but not for SIFT keypoints. The SIFT

keypoints inherently encode noise, so the additional noise did not result in further deterioration. If anything, the noise appears to actually help SIFT rejection in some cases like epipolar constraint. However, that could also be a result of parameter tuning for rejection metric thresholds.

In fact, all of the outlier rejection methods seemed highly sensitive to parameter tuning. For instance, a change in the *MLESAC E* tolerance from 35 to 37 px^2 resulted in a $\sim$22% decrease in outliers rejected for both landmarks with nominal noise. There are marginal performance changes between the methods, especially for SIFT with nominal noise.

4.2. Triangulation

Four triangulation methods were assessed: linear and nonlinear triangulation with and without pre-conditioning. The results are shown in Figures 10-13. Triangulation has the most uncertainty in the depth and linear triangulation without pre-conditioning is expected to perform the worst. Thus, depth error is assessed as the percent absolute change, which is defined here as,

$$\text{PAC} = 100\% \times \frac{|\bar{\delta}_z| - |\bar{\delta}_z^*|}{|\bar{\delta}_z^*|}. \tag{34}$$

Where $\bar{\delta}_z$ is the average triangulation depth error for a specific test case and $\bar{\delta}_z^*$ is $\bar{\delta}_z$ using linear triangulation without pre-conditioning. A PAC < 0 means the $\bar{\delta}_z$ was closer to 0 than the respective $\bar{\delta}_z^*$. Percent absolute change allows visualization of how much better or worse a method performed than basic linear triangulation.

The uncertainty in the x- and y-directions can be quantified with reprojection error,

$$\text{error}_{\text{rep}} = \|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|. \tag{35}$$

Here, $\boldsymbol{x}$ is the true image point and $\tilde{\boldsymbol{x}}$ is the image point reprojected from the triangulation 3D estimate.

There is an obvious performance jump in reprojection error from linear triangulation without pre-conditioning to any other method. That jump is not so apparent, and sometimes not even visible, for the depth PAC. The differences between the three other methods is almost indistinguishable. *Nonlinear PC* is a marginal winner in all cases, which was expected. However, the time and computation added on top of *linear PC* may not be worth the benefits.

One result that is not apparent in Figure 13 because of the PAC metric is that SIFT keypoints have a skew in the depth direction. The average depth error from triangulation for SIFT without system noise is shown in Figure 4. The errors are obviously not centered around zero. Figure 5 shows the same data for Eros face centers. A skew is slightly visible but it is not the same across all triangulation methods like it is for SIFT. Determining whether this skew is significant and repeatable requires further study. It could also be a factor in why the filter estimates in [25] would diverge.

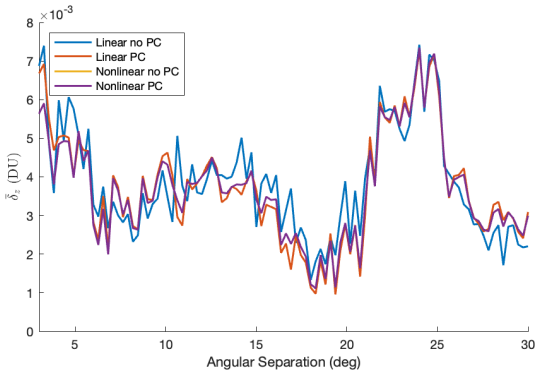

Figure 4. Triangulation average depth error for SIFT keypoints without any system noise.

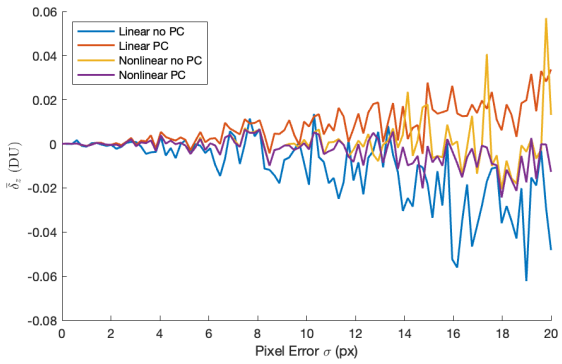

Figure 5. Triangulation average depth error for Eros face centers without any system noise.

4.3. Noise and Position

The influence of noise sources and relative position is assessed for five different metrics and the results are shown in Figures 14-17. The plots show both the average depth error per test case $\bar{\delta}_z$ and the reprojection error. All cases use nonlinear triangulation with pre-conditioning. Outliers in the SIFT keypoints are removed via ray-trace matching [5].

Distance of the camera to the asteroid center of mass (COM) appears to bare little affect on depth estimates. However, it is inversely related to the reprojection error in the presence of noise. As the camera moves farther from the surface, the landmarks become smaller and, thus, their centroids are more precise.

The relationship between triangualtion and angular separation is apparent in Figure 14. Interestingly, the angular separation requirements of SIFT and triangulation appear to cancel each other out in Figures 16 and 17.

Increasing pixel error has the obvious effect of increasing depth and reprojection error. However, the pixel error influence on depth gets drowned out by nominal noise and

error, even at $\sigma = 20$ px. This relationship is very similar to the influence of star tracker noise, but the star tracker influence is even less pronounced. This is an interesting result as it means reducing the error from pixel measurements and the star tracker has diminishing returns.

The camera position error has the greatest influence on depth and reprojection error. This affects camera projection matrix even more so than the rotation from ACAF to the camera frame because the rotation matrix does not induce scale changes.

5. Conclusions

Asteroids rendezvous missions are moving towards autonomy and lower size, weight, power, and cost avionics. This will improve mission turnaround time and decrease the barrier to entry for future missions. A potential option for such missions is to replace LiDAR and stereophotoclinometry with onboard, multi-agent stereovision using commercial off-the-shelf hardware. This paper explored the error and noise tolerances of the outlier rejection and triangulation sub-components of multi-agent stereovision. Two landmark definitions were evaluated: triangular mesh face centers and SIFT keypoint descriptors.

The noise and error sources analyzes were the pixel measurements, camera positions, and star tracker measurements. Pixel measurement error and star tracker measurements have a direct relationship with triangulation error. However, they both get washed out in the presence of nominal errors in a typical mission, especially camera position error. Camera position error has a significant influence on triangulation results. This means reducing pixel measurement and star tracker error has diminishing returns while reducing camera position error is paramount.

Two relative position metrics were also assessed: camera distance from the asteroid and angular separation between the cameras. The distance from the target had a marginal influence on triangulation so long as the asteroid was resolved in the image. On the other hand, triangulation error decreases as angular separation increases. Landmark correlation diminishes as angular separation increases. Thus, there is a sweet-spot of angular separation that returns the best triangulation results.

The epipolar constraint as well as maximum likelihood estimator random sample consensus using the fundamental matrix and essential matrix as models were tested as outlier rejection methods. These methods were also evaluated with and without point pre-conditioning. Overall, there were marginal differences between the results and each was highly sensitive to parameter tuning. In this case, the epipolar constraint is most likely a good choice as it does not require random sampling, singular value decomposition, and iteration to produce the same results.

Both linear and nonlinear triangulation were assessed, both with and without point pre-conditioning. As expected nonlinear triangulation with pre-conditioning resulted in the lowest reprojection and depth estimation error. However, linear triangulation with pre-condition and nonlinear triangulation without pre-conditioning were not too far behind. Thus, if a small amount of depth and reprojection error can be sacrificed, some computation time can be saved by using linear triangulation with point pre-conditioning.

One unexpected result of triangulation is that there was a bias in the depth error for SIFT keypoints. This was present for all triangulation methods and noise levels. The face centers showed a slight tendency toward a bias as measurement error was increased, but it was not as pronounced as in SIFT. This could be an interesting topic of future study as a bias in triangulation from the landmark choice can have repercussions in state estimation.

Nevertheless, there are many avenues of future work stemming from this paper. A natural continuation is to evaluate structure from motion techniques, many of which build on the mathematical principals discussed here. A computation time analysis of each algorithm is also an important next step. Other areas are increasing the number of agents or enabling bundle adjustment for multiple agents or sequential images. It would also be beneficial to swap out the asteroid for a man-made object (e.g. another spacecraft) or a planetary surface.

A note to the CS231A TA's: *The GitLab repository used for this work is part of a much larger project and private to members of the Space Rendezvous Laboratory. If you would like to see the repository, please send me an email at* `kdenn@stanford.edu`.

References

[1] E. Adams, D. O'Shaughnessy, M. Reinhart, J. John, E. Congdon, D. Gallagher, E. Abel, J. Atchison, Z. Fletcher, M. Chen, C. Heistand, P. Huang, E. Smith, D. Sibol, D. Bekker, and D. Carrelli. Double asteroid redirection test: The earth strikes back. In *2019 IEEE Aerospace Conference*, 2019.

[2] B. A. Archinal, M. F. A'Hearn, E. Bowell, A. Conrad, G. J. Consolmagno, R. Courtin, T. Fukushima, D. Hestroffer, J. L. Hilton, G. A. Krasinsky, G. Neumann, J. Oberst, P. K. Seidelmann, P. Stooke, D. J. Tholen, P. C. Thomas, and I. P. Williams. Report of the IAU working group on cartographic coordinates and rotational elements: 2009. *Celestial Mechanics and Dynamical Astronomy*, 109(2):101–135, 2011. ISBN: 0923-2958 1572-9478.

[3] C. Beder and R. Steffen. Determining an initial image pair for fixing the scale of a 3D reconstruction from an image sequence. In *28th Conference on Pattern Recognition*, pages 657–666, Sept. 2006.

[4] D. Capel, A. Fitzgibbon, P. Kovesi, T. Werner, Y. Wexler, and A. Zisserman. Matlab functions for multiple view geometry, 2012.

[5] K. Dennison and S. D'Amico. Comparing optical tracking techniques in distributed asteroid orbiter missions using ray-tracing. In *AAS/AIAA Space Flight Mechanics Meeting*, Charlotte, NC, Feb. 2021.

[6] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[7] R. W. Gaskell, O. S. Barnouin-Jha, D. J. Scheeres, A. S. Konopliv, T. Mukai, S. Abe, J. Saito, M. Ishiguro, T. Kubota, T. Hashimoto, J. Kawaguchi, M. Yoshikawa, K. Shirakawa, T. Kominato, N. Hirata, and H. Demura. Characterizing and navigating small bodies with imaging data. *Meteoritics & Planetary Science*, 43(6):1049–1061, June 2008.

[8] GOMSpace. NanoCam C1U Datasheet, 2016.

[9] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, Cambridge, 2004.

[10] R. I. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, Nov. 1997.

[11] P. S. Heckbert and M. Herf. Simulating soft shadows with graphics hardware. page 10, 1997.

[12] M. Hess-Flores, S. Recker, and K. I. Joy. Uncertainty, Baseline, and Noise Analysis for L1 Error-Based Multi-view Triangulation. In *2014 22nd International Conference on Pattern Recognition*, pages 4074–4079, Aug. 2014. ISSN: 1051-4651.

[13] Hongdong Li and R. Hartley. Five-point motion estimation made easy. In *18th International Conference on Pattern Recognition (ICPR'06)*, pages 630–633, Hong Kong, China, 2006. IEEE.

[14] A. S. Konopliv, J. K. Miller, W. M. Owen, D. K. Yeomans, J. D. Giorgini, R. Garmier, and J.-P. Barriot. A global solution for the gravity field, rotation, landmarks, and ephemeris of eros. *Icarus*, 160(2):289–299, 2002.

[15] Z. Kukelova, M. Bujnak, and T. Pajdla. Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *Procedings of the British Machine Vision Conference 2008*, pages 56.1–56.10, Leeds, 2008. British Machine Vision Association.

[16] R. Larson, R. P. Hostetler, B. H. Edwards, and D. E. Heyd. *Calculus with analytic geometry*. DC Heath Lexington, Massachusetts, USA, 1986.

[17] P. Lindstrom. Triangulation made easy. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1554–1561. IEEE, 2010.

[18] D. A. Lorenz, R. Olds, A. May, C. Mario, M. E. Perry, E. E. Palmer, and M. Daly. Lessons learned from OSIRIS-REx autonomous navigation using natural feature tracking. In *2017 IEEE Aerospace Conference*, pages 1–12, Big Sky, MT, USA, Mar. 2017. IEEE.

[19] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov. 2004.

[20] W. M. Owen, T. C. Wang, A. Harch, M. Bell, and C. Peterson. Near optical navigation at Eros. Quebec, Canada, July 2001. ISSN: 00653438.

[21] S. Palo, G. Stafford, and A. Hoskins. An agile multi-use nano star camera for constellation applications. In *AIAA/USU Conference on Small Satellites*, 2013.

[22] R. Park and A. Chamberlin. HORIZONS Web-Interface.

[23] Committee on the Planetary Science Decadel Survey, Space Studies Board, Division on Engineering and Physical Sciences, and National Research Council of the National Academies. *Vision into Voyages for Planetary Science in the Decade 2013-2022*. The National Academies Press, 2011.

[24] R. Siegwart, I. Nourbakhsh, and D. Scaramuzza. *Introduction to Autonomous Mobile Robots*. Massachusetts Institute of Technology, Cambridge, 2nd edition, 2004.

[25] N. Stacey, K. Dennison, and S. D'Amico. Autonomous asteroid characterization through nanosatellite swarming. In *IEEE Aerospace Conference*, Big Sky, Montana, Mar. 2022. IEEE.

[26] P. Torr and A. Zisserman. MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. *Computer Vision and Image Understanding*, 78(1):138–156, Apr. 2000.

[27] A. Vedaldi and B. Fulkerson. VLFeat: An Open and Portable Library of Computer Vision Algorithms, 2008.

[28] S. Watanabe, M. Hirabayashi, N. Hirata, N. Hirata, R. Noguchi, Y. Shimaki, H. Ikeda, E. Tatsumi, M. Yoshikawa, S. Kikuchi, H. Yabuta, T. Nakamura, S. Tachibana, Y. Ishihara, T. Morota, K. Kitazato, N. Sakatani, K. Matsumoto, K. Wada, H. Senshu, C. Honda, T. Michikami, H. Takeuchi, T. Kouyama, R. Honda, S. Kameda, T. Fuse, H. Miyamoto, G. Komatsu, S. Sugita, T. Okada, N. Namiki, M. Arakawa, M. Ishiguro, M. Abe, R. Gaskell, E. Palmer, O. S. Barnouin, P. Michel, A. S. French, J. W. McMahon, D. J. Scheeres, P. A. Abell, Y. Yamamoto, S. Tanaka, K. Shirai, M. Matsuoka, M. Yamada, Y. Yokota, H. Suzuki, K. Yoshioka, Y. Cho, S. Tanaka, N. Nishikawa, T. Sugiyama, H. Kikuchi, R. Hemmi, T. Yamaguchi, N. Ogawa, G. Ono, Y. Mimasu, K. Yoshikawa, T. Takahashi, Y. Takei, A. Fujii, C. Hirose, T. Iwata, M. Hayakawa, S. Hosoda, O. Mori, H. Sawada, T. Shimada, S. Soldini, H. Yano, R. Tsukizaki, M. Ozaki, Y. Iijima, K. Ogawa, M. Fujimoto, T.-M. Ho, A. Moussi, R. Jaumann, J.-P. Bibring, C. Krause, F. Terui, T. Saiki, S. Nakazawa, and Y. Tsuda. Hayabusa2 arrives at the carbonaceous asteroid 162173 Ryugu—A spinning top–shaped rubble pile. *Science*, 364(6437):268–272, Apr. 2019. Publisher: American Association for the Advancement of Science Section: Report.

[29] B. Williams, P. Antreasian, J. Bordi, E. Carranza, R. Chesley, C. Helfrich, J. Miller, W. Owen, and T. Wang. Navigation for near shoemaker: The first spacecraft to orbit an asteroid. 2001.

[30] D. K. Yeomans, J.-P. Barriot, D. W. Dunham, R. W. Farquhar, J. D. Giorgini, C. E. Helfrich, A. S. Konopliv, J. V. McAdams, J. K. Miller, W. M. Owen, D. J. Scheeres, S. P. Synnott, and B. G. Williams. Estimating the Mass of Asteroid 253 Mathilde from Tracking Data During the NEAR Flyby. *Science*, 278(5346):2106–2109, Dec. 1997.

[31] M. T. Zuber, D. E. Smith, A. F. Cheng, J. B. Garvin, O. Aharonson, T. D. Cole, P. J. Dunn, Y. Guo, F. G. Lemoine,

G. A. Neumann, et al. The shape of 433 eros from the near-shoemaker laser rangefinder. *Science*, 289(5487):2097–2101, 2000.

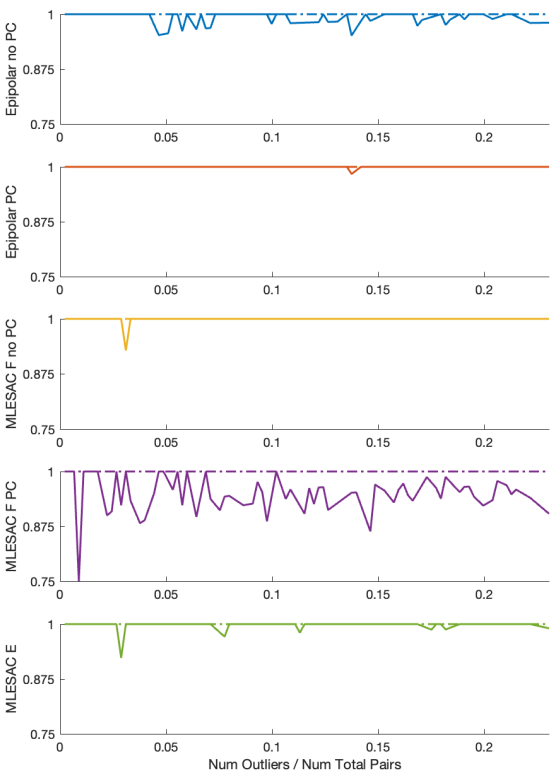

Figure 6. Outlier rejection results for Eros face centers without any system noise.

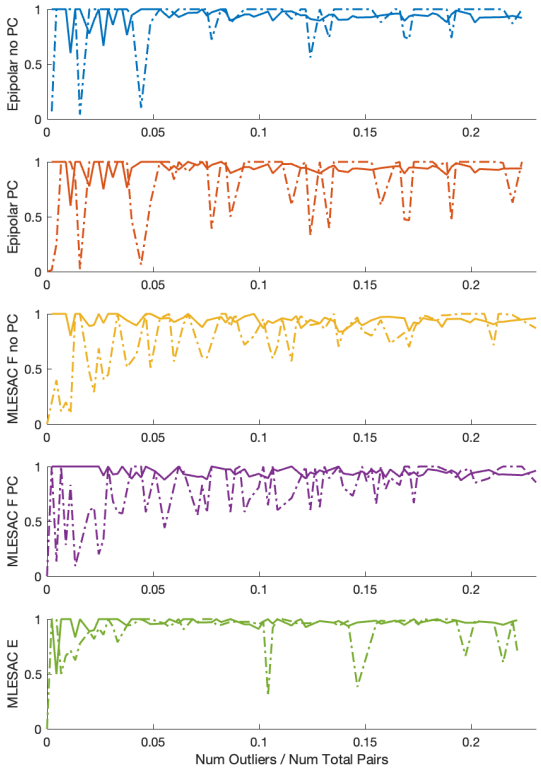

Figure 7. Outlier rejection results for Eros face centers with nominal noise applied.

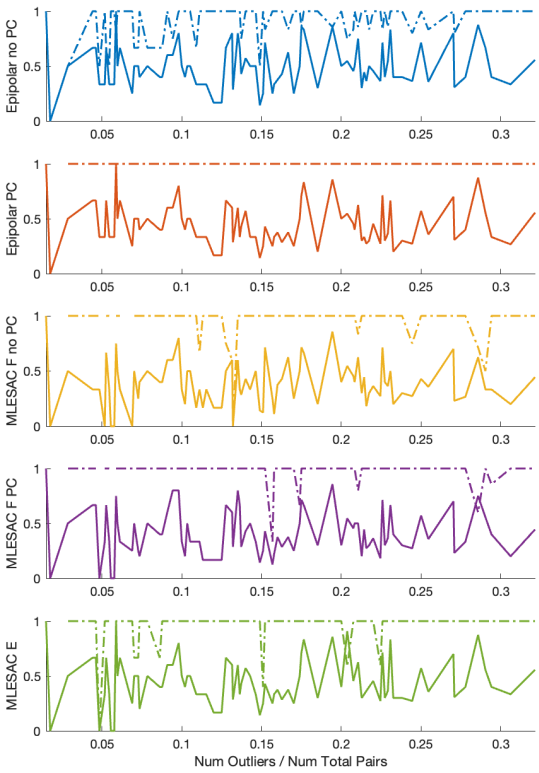

Figure 8. Outlier rejection results for SIFT keypoints without any system noise.

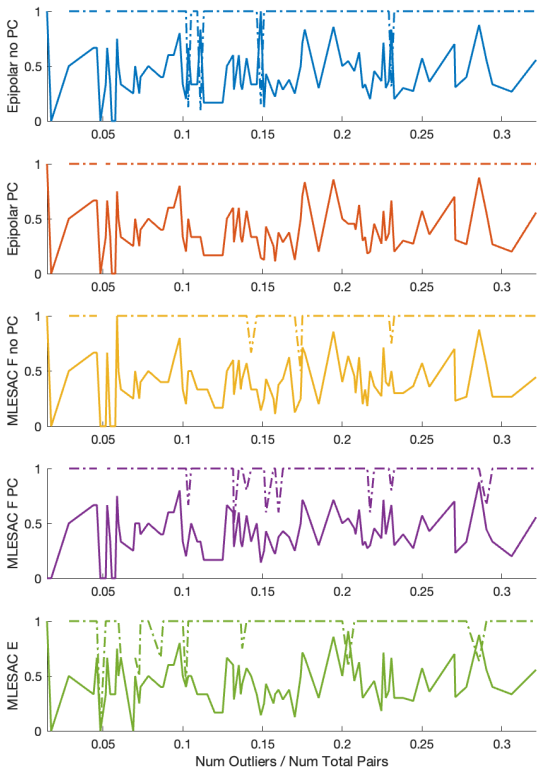

Figure 9. Outlier rejection results for SIFT keypoints with nominal noise applied.

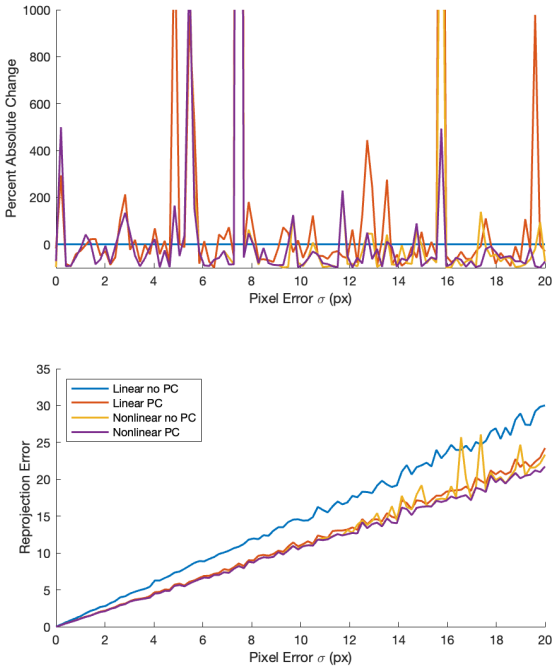

Figure 10. Triangulation results for Eros face centers without any system noise.

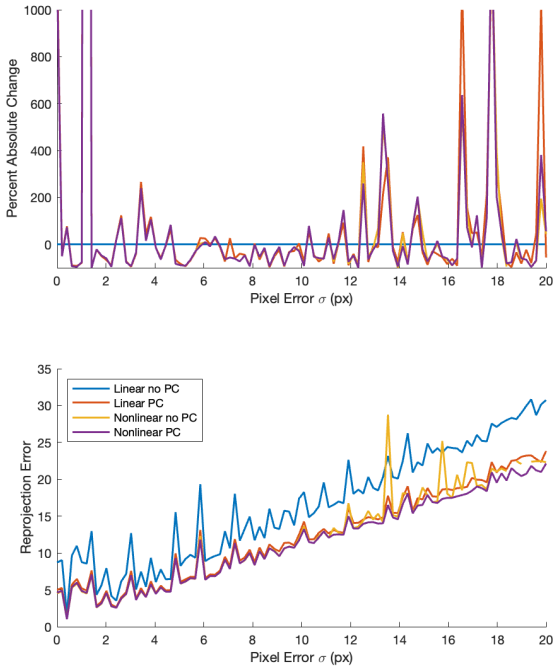

Figure 11. Triangulation results for Eros face centers with nominal noise applied.

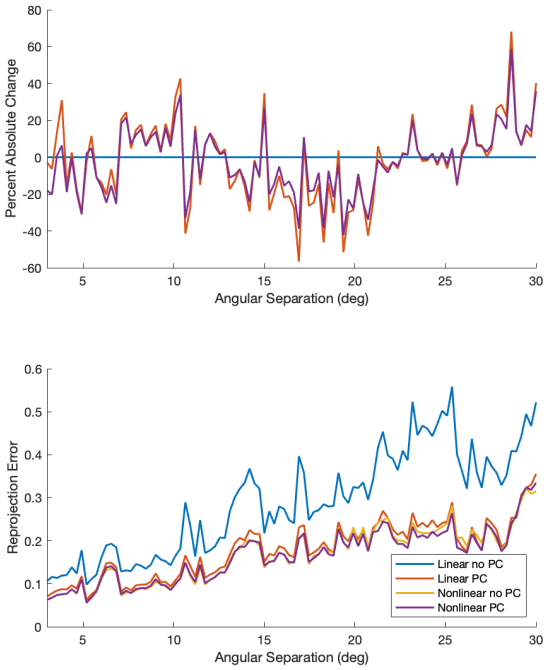

Figure 12. Triangulation results for SIFT keypoints without any system noise.

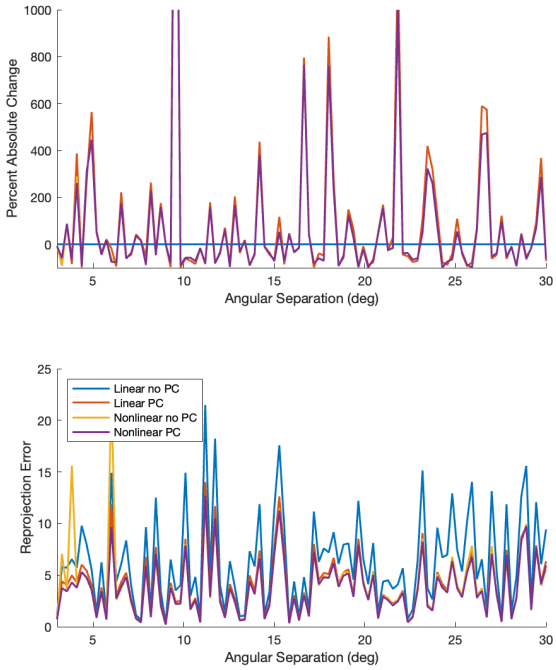

Figure 13. Triangulation results for SIFT keypoints with nominal noise applied.

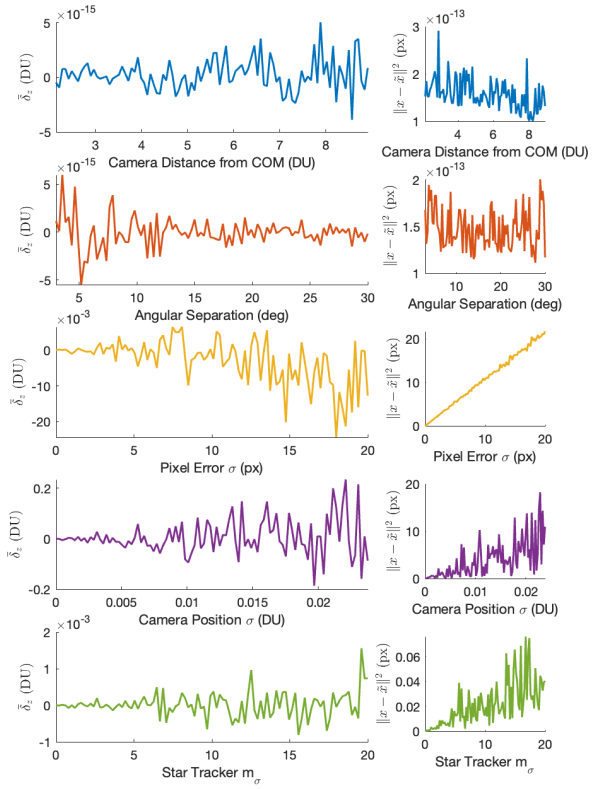

Figure 14. Position and noise results for Eros face centers without any system noise.

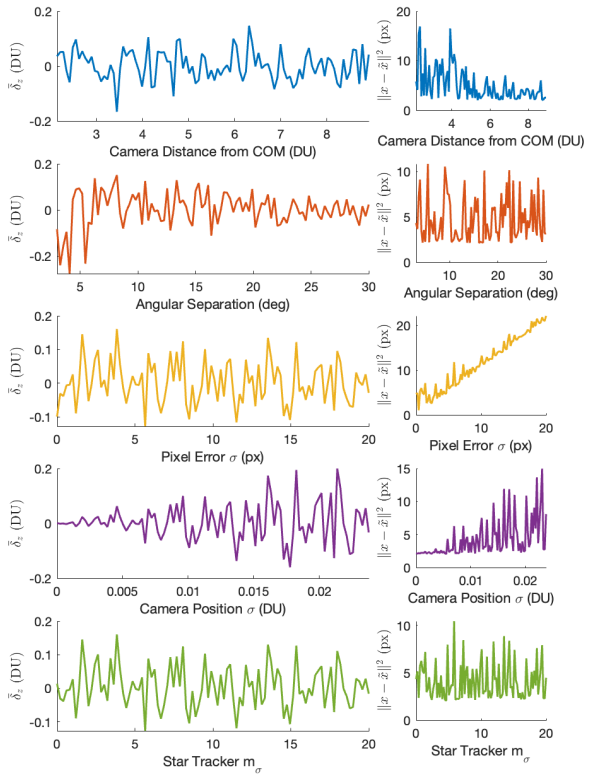

Figure 15. Position and noise results for Eros face centers with nominal noise applied.

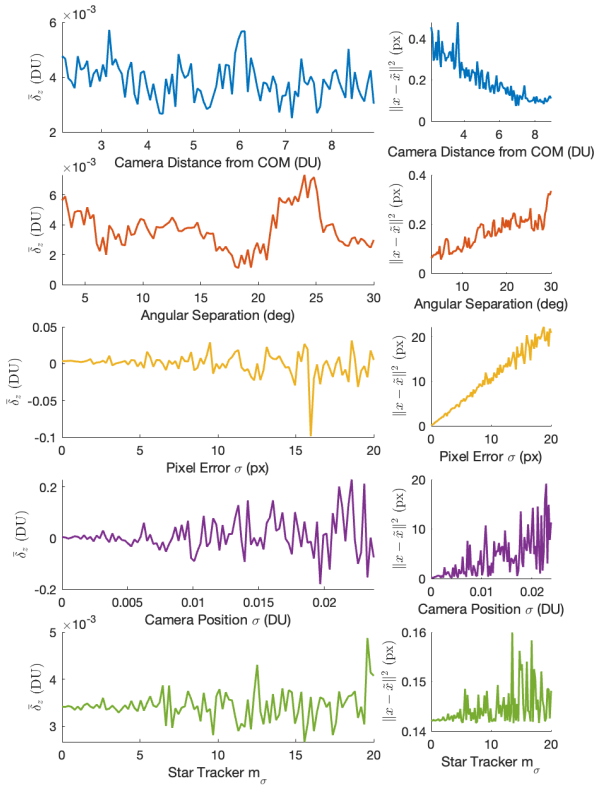

Figure 16. Position and noise results for SIFT keypoints without any system noise.

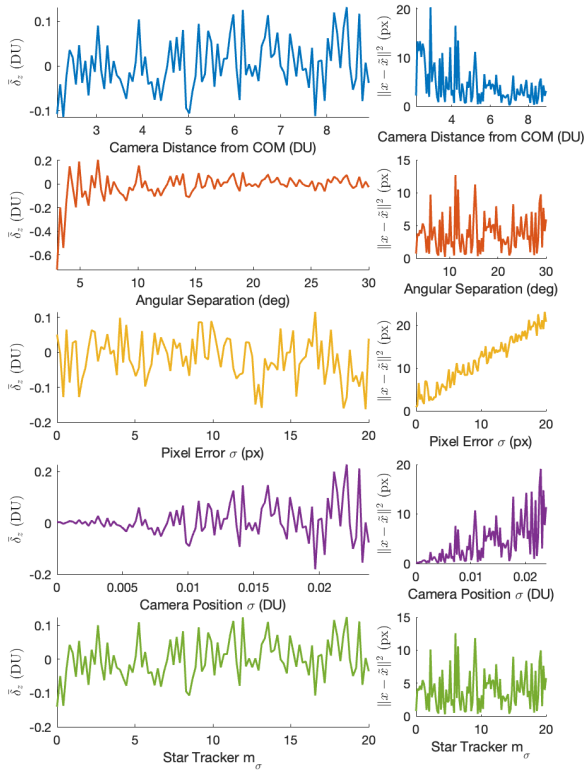

Figure 17. Position and noise results for SIFT keypoints with nominal noise applied.