IAC–16–C1.3.4

SPACECRAFT POSE ESTIMATION USING A MONOCULAR CAMERA

Jian-Feng Shi, Steve Ulrich
Carleton University, Canada, jianfeng.shi@carleton.ca, steve.ulrich@carleton.ca

Stéphane Ruel
Neptec Design Group Ltd, Canada, sruel@neptec.com

This paper propose a process for spacecraft pose estimation using monocular camera images. This process uses a unique combination of methods in image processing, feature extraction and matching, outlier removal via statistical consensus, and the use of perspective-n-point solvers including efficient-PnP and *SoftPOSIT* for camera to spacecraft pose estimation. While implementing the pose estimation process, two novel solutions were presented in the image processing and to enhance the iterative *SoftPOSIT* PnP solver. A histogram equalization by region scheme allows the adjustment of the pixel intensity with much faster speed than the conventional methods, this scheme handles images with pixel intensity already spanned to its full range. Secondly, three improvements were made to a perspective-n-point solver called *softPOSIT*. These improvements include techniques for reducing false matches as a result of local minimum trapping and strategies for iteration control parameter initialization by using the trace of the correspondence distance, and by using centroid matching. Results show the proposed pose estimation method is sufficiently stable and fast for real-time spacecraft pose estimation.

I. INTRODUCTION

With renewed interest for high density satellite constellations,[1] sample return mission from Mars,[2] spacecraft formation flying,[3] on-orbit servicing,[4,5] and much frequent proximity operations (ProxOps) between manned,[6,7] and unmanned autonomous space vehicles.[8–10] This motivates advancements in spacecraft ProxOps navigation technologies. In particular, the pose estimation of a spacecraft using a monocular camera for non-cooperative clients. Many technologies currently exist today for the tracking and pose determination for cooperative spacecraft clients. While using these devices, high degrees of accuracy can be resolved by systems such as NASA's space vision system (SVS)[11] which makes use of video markers, and ESA's ATV and JAXA HTV's use of RVS laser range finders (LRF),[12,13] NASA's advanced video guidance sensor (AVGS) can resolved targets at distance several kilometers away.[14] However, these systems require reflectors or markers to be placed on the client, and have limited range of operation where fiducial markers must be in view of the monitoring device. Some systems use LIDAR, such as DARPA's Orbital Express use a combination of photo, infrared camera, and LRF[15] and Neptec's TriDAR[16,17] do not require cooperative clients. However, the use of laser systems is more power intensive than photo or infrared cameras. The purpose of this study is to investigate software algorithm for spacecraft pose estimation using a monocular camera.

I.i Related Work

Earlier work on image based spacecraft pose estimation use *model-based* or *non-model* based approaches. Non-model based approach include the use of stereo egomotion estimation,[18] optic flow,[19] and structure from motion (SFM).[20] An example of this is by Augenstein and Rock[21] whom used the approach of scale invariant feature transform (SIFT) feature based simultaneous localization and mapping (SLAM)/recursive SFM for frame to frame tracking of the target spacecraft. Non-model based approach do not require knowledge of the target object but are challenged by the loss of particular tracked feature due to image condition or perspective projection. Another approach is model-based, where some knowledge the target shape is available. Model-based techniques include feature-based model tracking,[22] template matching,[23] contour tracking,[24] articulated object tracking,[25] and point correspondence.[26] Kelsey *et al*.[27] matched spacecraft model edge images to projected outlines from an internal wire frame model using reweighed least-squares iterations. Diaz and Abderrahim[28] and Arantes Jr.[29] extracted spacecraft point features through Hough Transform and used iterative model to image point correspondence method that are based in *Pose from Orthography and Scaling with ITerations* (POSIT). Miravet *et al*.[5] filtered the spacecraft from the background using pixel histogram and shape scoring and matched it to a projected model image. Alternatively, Sharma and D'Amico[30] use points from edge features as inputs for various perspective-n-point(PnP) solver. Chen *et al*. demonstrated six degrees of freedom (DOF) pose estimation with a moving camera and target by using a combination of image feature, efficient PnP (ePnP) and Kalman filtering.[31]

Inspired by the aforementioned authors, this investigation is based on using image features with descriptors and solving the PnP problem. Uniquely, rather than matching the internal 3D model directly onto the extracted feature. A homography matrix is computed to map the model projection on to the image plane. This approach is more stable and is without image outliers when compared to the direct matching approach. This paper is organized as follows, Sec. 2 shall provide the pose estimation process, including 3D modeling, image processing, feature extraction and matching, homography and outlier rejection, and PnP solvers *SoftPOSIT* and ePnP. Sec. 3 provides the pose estimation simulation setup. Sec. 4 provides pose estimation results. Finally, Sec. 5 concludes this study.

II. POSE ESTIMATION

This study propose a method for pose estimation that consists of two main sections. First, there is a feature matching phase to obtain homography between the training image and the query image. Second, the algorithm performs a projection of the internal 3D model and performs a rotation of the projected image point based on the homography matrix from step one. The rotated image projection is used by the PnP solver along with the internal 3D model to determine the object pose from the camera. It was found by earlier work line and corner features are not stable to be used for model point correspondence. Lighting and motion conditions resulted in unpredictable changes in rudimentary features such as lines and corners. Significant time was spent on image processing to obtain consistent line and corner feature inputs. The initial approach was non-sustainable for real-time (RT) operations as image processing tasks can take enormous computational resources. The severity of the aforementioned issue will only grow with the increasing camera resolution. By contrast, advanced image features including feature descriptors are designed to be invariant to translation, rotation, scale, and to some extent illumination changes.[32] Experiments also have shown the use of these image features to be more stable when compared to rudimentary features. Furthermore, when directly matching to rudimentary features, there is a high number of outliers from both the image and from model points that should be occulted. Instead of investing run time to process geometric algorithms to occult model points based on perspective views, it is better to use the model points directly. Consequently, this approach also eliminated the outlier issue entirely. The following sections describes the details of the entire pose estimation process.

II.i Pose Estimation Process

The pose estimation process flow is shown in Fig. 1. The image is first acquired by video using either a virtual camera in a 3D environment or a real world camera. A training image is initially taken, either on orbit or by ground test. This training image should have the most comparable contrast gradient to the operational image for best feature match. Once processed by the feature extraction algorithm, only the feature keypoints and descriptors are stored while the training image itself can be discarded. Features of multiple training images taken from different pose and illumination conditions can be stored in an on-board library for later processing. During the operation phase, the query image is fed into the image processing function. This function is responsible for quickly restoring the image and extracting image features and descriptors, it uses the internally stored test image features for feature matching. The output of this function is a set of matched feature points between the training image and the query image. The next function is the projection, RANSAC, and stabilization block. This block computes the image homography using the matched features. A RANdom SAmple Consensus (RANSAC)[33] algorithm is used to test the feature matches for outlier points removal. Next, the internal model is projected to the image plane using the calibrated camera intrinsic properties and transformed using the computed homography matrix. A stabilization sentinel function is additionally used to remove invalid homography transformations. Next, transformed image point projections and the body coordinate 3D model points are sent to the PnP solver for camera to body frame pose. Finally, the computed position and orientation is processed through spacecraft kinematic calculator for target pose output in the desired coordinate. Additional motion sensor information local to the monitoring tracker spacecraft can be used in this block to separate absolute pose and motion of the target spacecraft.

II.ii 3D Modeling

The internal model consists of stored vertices of elementary shapes such as rectangles, cylinders, ellipses and cones shown in Fig. 2. The elementary shapes can be used to construct more complex objects such as various spacecraft shown in Fig. 3. The inherit feature of the proposed pose estimation approach is the ability of using very simple shapes in representing the target spacecraft. This alleviate the need to model reduce complex CAD models, which may remove important feature information.

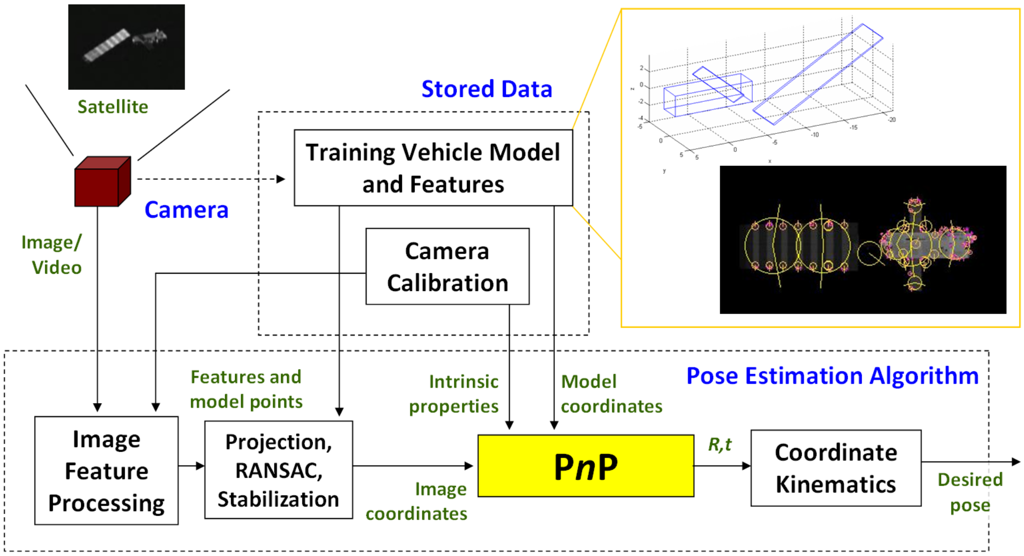


Fig. 1: Pose estimation process.

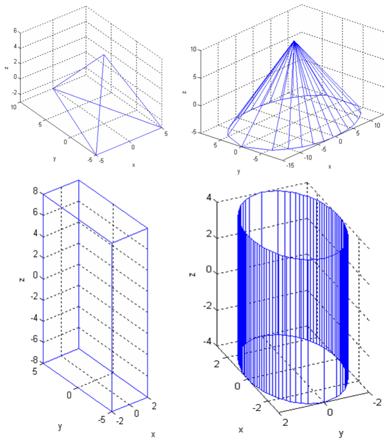


Fig. 2: Elementary building shapes.

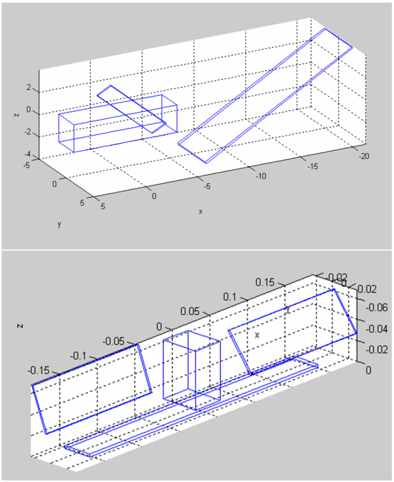


Fig. 3: Complex spacecraft models.(Top:Envisat,Bottom:Radarsat)

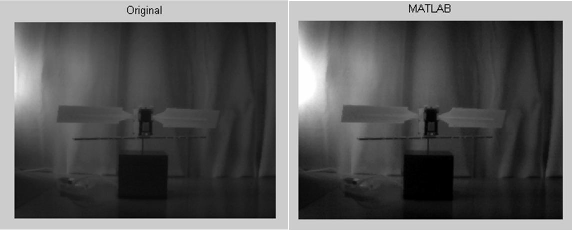


Fig. 4: Histogram expanded IR image. (Left:Original,Right:HE applied)

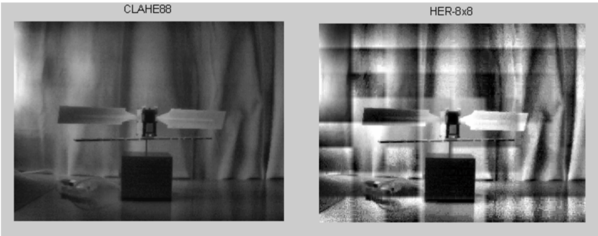


Fig. 5: Processed IR image.(Left:CLAHE,Right:HER)

Fig. 6: Extreme illumination IR image.(Left:Original,Right:HER)

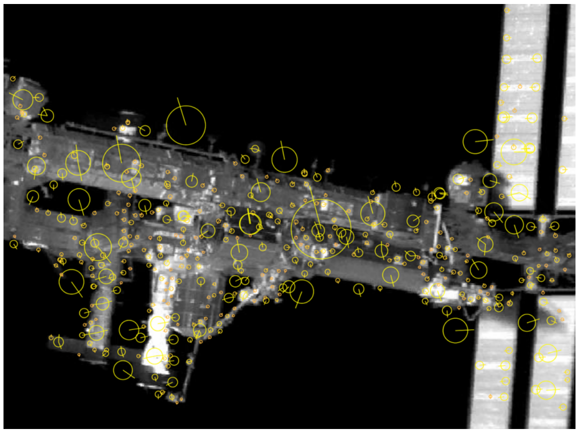


Fig. 7: SIFT feature of ISS IR image. Courtesy NEPTEC

II.iii Image Processing

The first image processing task is to clean and restore the input image. The fastest, simplest, and arguably the most effective image processing technique is histogram equalization (HE). However, space imagery is known for spanning over large illumination ranges. An example of a lab environment infrared (IR) image demonstrates this problem as shown in Fig. 4 A widely used improvement to HE is called contrast limited adaptive histogram equalization (CLAHE).[34] Although this technique has been successfully applied to space imagery,[35] it is however costly to run. A simpler technique using much less computation resource was developed as part of this pose estimation software. The idea is to divide up the image and perform HE on each region, this is named HE by region (HER). The results of HER and CLAHE of the original image is shown in Fig. 5 By comparison, the CLAHE (8x8) image uses 61.6ms while the HER (8x8) image uses 4.8ms. [a]. To relieve sharp changes across region, a mixing zone was created to smooth the boundary. This mixing zone can be design to be transparent to edge detection processes as shown in Fig. 6.

II.iv Feature Extraction

Reliable target detection and matching is the most important task in the pose estimation process. It is therefore crucial to use stable and reliable image features for this task. Modern image features that are invariant to translation, rotation, and scale was first introduced by Lowe in 2004 called scale-invariant feature transform (SIFT).[32] Instead of only finding key points, the SIFT feature also include a 128 dimension descriptor vector that allow for more precise matching. Many variations have been developed to improve computation speed, such as speeded-up robust features (SURF)[36] and the latest Accelerated-KAZE (AKAZE)[37] but none have surpass SIFT in its accuracy and reliability. [*] Figure 7 shows the SIFT features extracted from an ISS IR image. The downside of the SIFT feature is it requires higher processing than the newly developed features and it is a commercially licensed technology. The latest trend in image features development is by using binary descriptors. Binary descriptors requires much less memory storage, it can take advantage of XOR operators and certain processor architectures. The first kind of binary descriptor is the binary robust independent elementary features (BRIEF).[38] It uses randomly generated binary patches to achieve the same Difference of Gaussian (DoG) as the SIFT feature. BRIEF descriptors are extracted from Shi-Tomasi corners.[39] The BRIEF random patch is shown in Fig. 8 The weakness in BRIEF is that it is not rotationally invariant. This weak-

---

[a] Simulated in 32bit-Windows Matlab with 2.4GHz Intel® Core™ 2 Quad Q6600 processor

[*] The SIFT software was kindly provided by Dr. Lowe of the University of British Columbia for this academic research purpose.

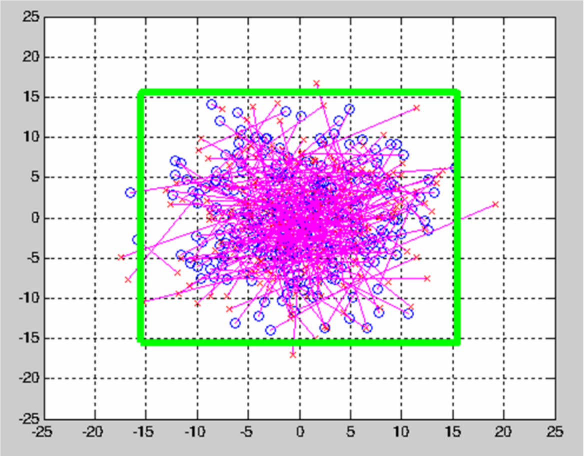


Fig. 8: BRIEF Gaussian binary patch.

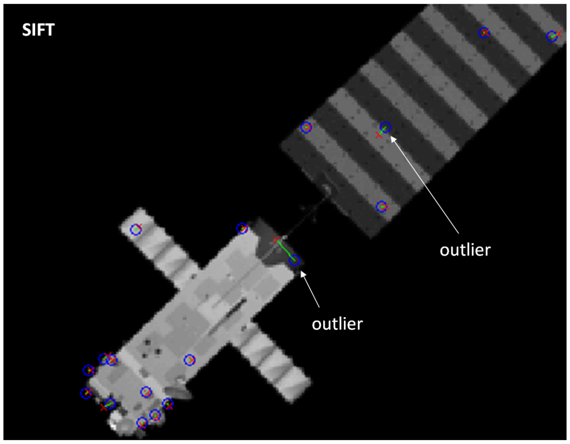


Fig. 9: SIFT feature matching outliers.

ness can be improved by using Oriented-fast and Rotated BRIEF (ORB)[40] keypoint and descriptor. Where orientation of the corner feature can be determined using the *intensity centroid*. Without rotation invariant accommodation, the BRIEF descriptor was tested to be resilient up-to 15 degrees target rotation. For frame to frame pose tracking, it is not required to have large angle rotational invariant descriptors that uses higher computational resource.

Feature Matching

Feature matching of the descriptor can be done using k-Nearest Neighbour methods.[41] Specifically, two nearest neighbour distances are compared to eliminate outlier matches. Lowe[32] suggest the true match should have a small first match to second match distance ratio since there can only be one match to the image feature. This match ratio is tested to be 0.6. The same idea can extend to the polar domain. Where angles to first and second match is compared instead of distance. The angle approach reduced the ratio comparison to one dimension and is by order of magnitude faster. However, it is less accurate than the distance comparison.

II.v RANSAC Homography and Sentinal

Standard method[42] in homography was used to obtain transformation between the training image and the query image to satisfy the following image coordinate relationship,

$$\tilde{\mathbf{p}}_q = \mathbf{H}\tilde{\mathbf{p}}_t \qquad [1]$$

where $\mathbf{H}$ is the homography matrix, and $\tilde{\mathbf{p}}$ is the homogeneous image feature coordinate for the training (t) and the query (q) image. It is important to note homography is not guaranteed in general. Future development of the pose estimation algorithm will focus on supplementing this process block so it can be robust under all conditions.

The RANdom Sample and Consensus (RANSC) algorithm is used to remove outlier nearest neighbour matches.[33] The number of RANSAC trials is computed using the following equation,

$$T = \frac{log(1-p)}{log(1-(1-e)^s)} \qquad [2]$$

where T is the number of trials, p is the desired success probability, e is the outlier ratio, s is the number of sampled points. Based on experimental testing, the typical outlier using 0.6 matching ratio is 5% while using a sample of 20 points to achieve 90% success probability requires 5 trials. This only amounts to $2ms$ computation time and do not affect the RT capability of the software. RANSAC outliers are shown in Fig. 9

To handle invalid homography transformation, a sentinel checker is added to prevent bad image points to be passed to the PnP solver. The sentinel checker evaluates the maximum movement of the model boundary points. An assumption is made such that the target nor the tracker have realistic velocities that would exceed some allowed pixel movement range within one frame. Given the short time between one frame, the pose is held constant for the next frame's data. The movement range adaptively increase as more bad frames are returned. Upon receiving a good frame, the movement range is restored to the original value.

II.vi Perspective-n-Point

The engine of the pose estimation method is the use of a PnP solver, the goal is to solve the target body pose relative to the camera given point coordinates local to the body frame and the projected image points. Solutions to the PnP problem can be classified by *iterative* and *non-iterative* methods. Iterative approach generally minimizes some energy function to match the estimated projected image point to the true image. Earlier iterative methods include O(n) Direct Linear Transformation (DLT)[43] and O(nm) *SoftPOSIT*.[26] A more detailed description for *SoftPOSIT* will follow in Sec. 2.6.2. The LHM iterative solver by Lu *et al.*[44] is widely regarded as highly accurate for the PnP solution. Schweighofer *et al.*[45] propose a O(n) global optimal solution using the SeDuMi semi definite positive program (SDP) solver. The state-of-the-art iterative PnP solvers includes works by Fan *et al.*,[46] SoftSI by Zhou *et al.*,[47] and method related to Scaled Orthographic Projection (SOP) by Sun *et al.*[48] Finally, the latest iterative PnP method by Urban *et al.*[49] considers maximum likelihood PnP (MLPnP) solutions to image observation uncertainties.

Close form non-iterative solution include solving for 3 points (P3P)[50,51] or 4 points (P4P).[33] by geometric or algebraic closed form formulas. Lepetit *et al.*[52,53] is the first to introduced an O(n) close form solution called efficient-PnP (ePnP). This methods uses four virtual control points (VCP) to span any number of the model correspondence points (CP). It then uses the inter-VCP geometric distance constraints to solve for the VCP in the camera frame. Further details for the ePnP method is provided in Sec. 2.6.3. Gao *et al.*[54] and Ferraz *et al.*[55] further improved upon the ePnP method by adding iteration on the depth matrix and outlier rejection respectively. The Robust PnP (RPnP)[56] solution retrieves the roots of a seventh order polynomial resulting from least square minimization of P3P problems. Zheng *et al.*[57,58] propose two direct minimization methods using Gröbner basis solver resulting in Accurate and Scalable PnP (ASPnP) and Optimal PnP (OPnP). Kneip *et al.*[59] and Nakano[60] both proposes extensions to the direct lease squares (DLS)[61] method to include non-central camera rays and by global optimization with Cayley parameterization respectively.

Given the variety of various PnP solutions, two elementary solvers, an iterative and a non-iterative, were selected to serve as initial baseline. The two solvers are *softPOSIT*[26] and ePnP.[53] Performance of each solver its strength and weakness were evaluated. Some improvements were made to the *SoftPOSIT* solver.[62] Further discussion of the improvements and methods for each solver are provided in the following sections.

Camera Model

This section provides problem definition for the various PnP techniques. Define a tracker body $\mathcal{F}_{SB}$ equipped with a single camera $\mathcal{F}_{VW}$ pointed towards a target object $\mathcal{F}_{CB}$. The frames $\mathcal{F}_{SB}$ and $\mathcal{F}_{CB}$ are located at the object Center of Geometry (COG). The camera frame $\mathcal{F}_{VW}$ has its z axis pointed outwards from the boresight of the camera, with the y axis pointed vertically downwards and x completing the right-handed frame. The frame $\mathcal{F}_{ia}$ is centered on the camera image with its x and y axes being parallel to $\mathcal{F}_{VW}$. Figure 10 shows the various coordinate systems considered. Let define PW_k be the k^{th} point on the target that is projected on the image plane. This projected point is denoted as pi_k. The total number of points from the target model is M. These points are the corner vertices of the target body. The objective is to determine the pose, *i.e.* , $\mathbf{t}$ and $\mathbf{R}$ of the target with respect to the camera defined as follows,

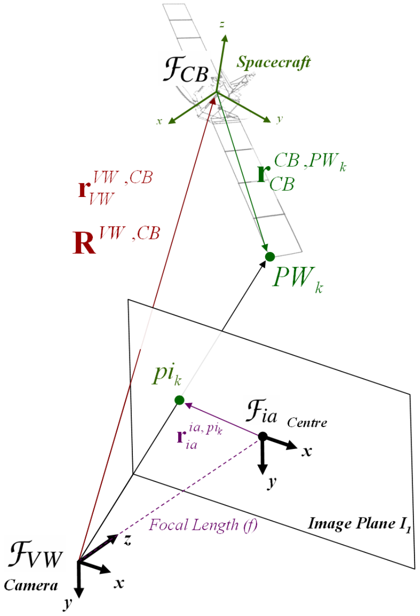


Fig. 10: Camera coordinate system definition.

$$\mathbf{t} = \mathbf{r}_{VW}^{VW,CB} = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T \quad [3]$$

$$\mathbf{R} = \mathbf{R}^{VW,CB} = \begin{bmatrix} \mathbf{R}_x^T \\ \mathbf{R}_y^T \\ \mathbf{R}_z^T \end{bmatrix} \quad [4]$$

where $\mathbf{t}$ is the position of $\mathcal{F}_{CB}$ w.r.t. $\mathcal{F}_{VW}$ expressed in $\mathcal{F}_{VW}$ and $\mathbf{R}$ is the rotation matrix rotating $\mathcal{F}_{CB}$ to $\mathcal{F}_{VW}$, and $\mathbf{R}_x^T$, $\mathbf{R}_y^T$, and $\mathbf{R}_z^T$ are row matrices of $\mathbf{R}$. These are also the $\mathcal{F}_{VW}$ unit vectors expressed in $\mathcal{F}_{CB}$. Let define N be the number of detected image points Let define the model points relative to and expressed in the body frame be denoted by $\mathbf{P}_k^w = \mathbf{r}_{CB}^{CB,PW_k}$, where $k = 1 \ldots M$. Let define the model points observed and expressed by the camera frame be $\mathbf{P}_k^c = \mathbf{r}_{VW}^{VW,PW_k}$. The image point and CP in homogeneous coordinates are defined by $\tilde{\mathbf{p}}_j = \begin{bmatrix} \mathbf{r}_{ia}^{ia,pi_j T} & 1 \end{bmatrix}^T = \begin{bmatrix} u_j & v_j & 1 \end{bmatrix}^T$ and

$\tilde{\mathbf{P}}^w_k = \left[\begin{array}{cc} \mathbf{P}^{wT}_k & 1 \end{array} \right]^T$ respectively. The standard camera intrinsic properties are denoted by the following,

$$\mathbf{K} = \left[\begin{array}{ccc} fS_x & 0 & \tilde{o}_x \\ 0 & fS_y & \tilde{o}_y \\ 0 & 0 & 1 \end{array} \right] \qquad [5]$$

where, f is the focal length, S is the pixel scale conversion, and $\tilde{o}$ is the pixel distance from the image coordinate origin to the image center. The standard camera equation is,

$$Z^c_k \tilde{\mathbf{p}}_k = \mathbf{K} \left[\begin{array}{cc} \mathbf{R} & \mathbf{t} \end{array} \right] \tilde{\mathbf{P}}^w_k \qquad [6]$$

where Z^c_k is the z component of $\mathbf{P}^c_k$.

SoftPOSIT

David *et al.*[26] solve the PnP problem by combining Simulated Annealing (SA)[63] and Scaled Orthographic Projection (SOP).[64] This method is termed *SoftPOSIT*, it is a combination of *Softassign*[65,66] and *Pose from Orthography and Scaling with ITerations*. *SoftPOSIT* is an iterative point correspondence scheme minimizing a global energy function based on points differences from weak projection. There has been several implementations of the *SoftPOSIT* method in terrestrial applications. For example, Jager *et al.*[67] used *SoftPOSIT* to determine the pose of a ground vehicle based on thermal camera images, and previously mentioned Diaz and Abderrahim used *SoftPOSIT* for spacecraft pose estimation.[68] Once converged, *SoftPOSIT* can produce fast and accurate pose estimation results; however, some shortcomings of the algorithm include local minimum trapping and iteration control parameter to correspondence compatibility. Enhancements for these issues are further discussed in later sections. Let define the *scaling ratio* s as $^f/_{t_z}$ and the *prospective ratio* w_k as $^{Z^c}/_{t_z}$, such that the prospective scaling term $^f/_{Z^c}$ can be replaced with $^s/_{w_k}$. If the object is far away from the camera and *field of view* (FOV) is not abnormally large, then w_k is nearly one. w_k can be written as

$$w_k = \frac{\mathbf{R}^T_z \mathbf{P}^w_k}{t_z} + 1 \qquad [7]$$

Let define the pose matrices $\mathbf{Q}_x \triangleq s \left[\begin{array}{cc} \mathbf{R}^T_x & t_x \end{array} \right]^T$, $\mathbf{Q}_y \triangleq s \left[\begin{array}{cc} \mathbf{R}^T_y & t_y \end{array} \right]^T$. The distance between the projected model points and the camera points is

$$d^2_{jk} = (\mathbf{Q}^T_x \tilde{\mathbf{P}}^w_k - w_k u_j)^2 + (\mathbf{Q}^T_y \tilde{\mathbf{P}}^w_k - w_k v_j)^2 \qquad [8]$$

A *Global Objective Function* is formulated as:

$$E = \sum_{j=1}^{N} \sum_{k=1}^{M} m_{jk} (d^2_{jk} - \alpha) \qquad [9]$$

For a maximum correspondence between image and model points, the partial derivative of E with respect to the pose matrices is set to zero. where the weights m_{jk} are computed at every step based on SA principals as follows,

$$m_{jk} = \gamma \exp(-\beta(d^2_{jk} - \alpha)) \qquad [10]$$

where γ is a normalization factor and α is to allow amplification of the d_{jk} distance. The corresponding distance between image and model points can be grouped in a matrix format as following:

$$\mathbf{D} = \left[\begin{array}{ccc} d^2_{1,1} & \cdots & d^2_{1,M} \\ \vdots & \ddots & \vdots \\ d^2_{N,1} & \cdots & d^2_{N,M} \end{array} \right] \qquad [11]$$

A major decrease in the *SoftPOSIT* estimation accuracy is caused by local minimum trapping[26,67,68] where further iterations cannot force the model pose away from the local minimum. Previous work[62] have shown a precomputation phase prior to full *SoftPOSIT* iteration can bring the solution away from the local minimum. To select the likely global optimal path, *the smallest maximum of the closest image to model correspondence for all n initial orientations is the likely path towards the global minimum.* The above statement can be evaluated by the following steps. First, write the correspondence matrix from Eq. (11) as an array of column matrices $\mathbf{D} = \left[\begin{array}{ccc} \mathbf{D}_1 & \ldots & \mathbf{D}_M \end{array} \right]$, these columns represent the correspondence from all the detected image points to one model projected point. The minimum value in the k^{th} column represent the closest image to model correspondence for the k^{th} model point, that is,

$$\hat{D}_{kw} = \underset{j \in N}{\operatorname{argmin}} \{ D_{jkw} \} \qquad [12]$$

The maximum of these represent the *goodness* as the result from optimizing from the w^{th} initial orientation.

$$\check{D}_w = \underset{k \in M}{\operatorname{argmax}} \{ \hat{D}_{kw} \} \qquad [13]$$

Finally, the minimum $\check{D}_w$ signals the most likely path to the global minimum solution, as,

$$\hat{D}_g \triangleq \underset{w \in n}{\operatorname{argmin}} \{ \check{D}_w \} \qquad [14]$$

In the spirit of annealing optimization, this precomputation phase is referred to as *preheating*.

From Eq. (10), the iteration control parameter β is used to control the annealing process. When β is low, the weighting will be high and vise-versa. The determination for β is largely based on trials and errors. Selecting the iteration control parameter can be difficult as improper matching between β and the correspondence matrix $\mathbf{D}$ may lead to numerical anomalies. In view of this, a simple β initialization is herein proposed as,

$$\beta_0 = F\frac{(M+N)/2}{\mathrm{tr}(\mathbf{D})} \qquad [15]$$

where F is some scaling constant to be set to 2 as the tested optimal value.

The exponential term in Eq. (10) requires an appropriate initial temperature to be selected in order to prevent the weighting m_{jk} from becoming zero or infinity due to numerical issues. The proposed method produces a iteration control parameter β that is compatible with the point correspondence, and eliminates the need for tedious trial and error tuning campaigns of β. Specifically, it uses the image point cloud centroid to estimate the target COG. This is done by extracting $\mathbf{t}$ terms from the optimal minimum $\mathbf{Q}$ and solving by Newton's method.[62]

Efficient PnP

The efficient-PnP (ePnP) method was initially developed by Moreno-Noguer[52] and Lepetit[53] and has been widely used as a fast non-iterative scheme for solving the PnP problem. This method uses virtual control points (VCP) as basis for all the model correspondence points (CP), and provide bounding conditions by using geometric restrictions from inter-VCP distances. For better accuracy, a Gauss-Newton iteration is added at the end to converge the solution. The computation procedure for the ePnP solution are provided as follows. For stable solution at least six points is preferred for number of correspondence, further discussion on the optimal number of CP to be selected from the spacecraft model will follow in Sec. 4. The centroid of the CP set is

$$\bar{\mathbf{P}}^w = \frac{1}{M}\sum_{k=1}^{M}\mathbf{P}_k^w \qquad [16]$$

The distance of the CP to the centroid is $\mathbf{d}_k^w = \mathbf{P}_k^w - \bar{\mathbf{P}}^w$. The first VCP is the computed centroid, the rest of the VCP are computed based on principal component analysis (PCA) as follows. Let $\mathbf{X}$ contain all M columns of $\mathbf{d}_k^w$ The Eigen vector and Eigen values of $\mathbf{X}$'s covariance matrix is,

$$\begin{aligned}\mathbf{S} &= \mathbf{X}\mathbf{X}^T \\ &= \mathbf{P}_s\mathbf{\Lambda}_s\mathbf{P}_s^T\end{aligned} \qquad [17]$$

where $\mathbf{e}_i$ are the Eigen vector columns of $\mathbf{P}_s$, and the diagonal values of $\mathbf{\Lambda}_s$ are the Eigen values λ_i^2 for $i = 1\ldots3$. The rest of the VCP are defined as,

$$\mathbf{c}_{i+1} = \mathbf{c}_1 + \sqrt{\frac{\lambda_i^2}{M}}\mathbf{e}_i \qquad [18]$$

Each CP is then the barycenter of the VCP and can be express as,

$$\mathbf{p} = \sum_{i=1}^{4}\alpha_i\mathbf{c}_i \qquad [19]$$

The α coefficients can be obtained by computing the following,

$$\alpha_1 = 1 - \sum_{i=2}^{4}\alpha_i \qquad [20]$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{c}_2 - \mathbf{c}_1 & \mathbf{c}_3 - \mathbf{c}_1 & \mathbf{c}_4 - \mathbf{c}_1 \end{bmatrix} \qquad [21]$$

$$\begin{bmatrix}\alpha_2\\ \alpha_3\\ \alpha_4\end{bmatrix} = \mathbf{A}^{-1}(\mathbf{x} - \mathbf{c}_1) \qquad [22]$$

Next, the camera intrinsic properties, the projected image points are used with the VCP to relate the VCP from the camera frame and the body frame. It can be shown the α coefficients are the same for CP relative to the body and the camera frame. This relationship allows geometric restrictions to be equated from both frames and ultimately return the pose of the body frame relative to the camera frame. Substituting the VCP definition from Eq. 19 into Eq. 6, let $\mathbf{x}^{cT} = \begin{bmatrix} \mathbf{c}_1^{cT} & \mathbf{c}_2^{cT} & \mathbf{c}_3^{cT} & \mathbf{c}_4^{cT} \end{bmatrix}$ where c denotes relative to and express in the camera frame,

$$\mathbf{M}_k = \alpha_k \otimes \begin{bmatrix} fS_x & 0 & \tilde{o}_x - u_k \\ 0 & fS_y & \tilde{o}_y - v_k \end{bmatrix} \qquad [23]$$

where the $\mathbf{M}$ matrix consists of $2M$ rows from the $\mathbf{M}_k$ matrix from Eq. 23. $\otimes$ is the Kronecker product. Then the following equation can be derived.

$$\mathbf{M}\mathbf{x}^c = \mathbf{0} \qquad [24]$$

Computing the singular value decomposition (SVD) of $\mathbf{M}\mathbf{M}^T$,

$$\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{M}\mathbf{M}^T \qquad [25]$$

$\mathbf{x}^c$ is the linear combination of the last four singular vectors of $\mathbf{U}$ as follows,

$$\mathbf{x}^c = \sum_{l=1}^{4}\beta_l\mathbf{v}_l \qquad [26]$$

Lepetit[53] found there can be four solution scenarios as $l = 1\ldots4$ and the order is proportional the the camera focal length. Experience shown the most stable solution results from the one singular value scenario. To solve for the coefficient β, let $\mathbf{v}_l^T = \begin{bmatrix} \mathbf{v}_{l1}^T & \mathbf{v}_{l2}^T & \mathbf{v}_{l3}^T & \mathbf{v}_{l4}^T \end{bmatrix}$. Six unique geometric distances between the body VCP can be found

$$\rho_r = \|\mathbf{c}_a^w - \mathbf{c}_b^w\|^2 \qquad [27]$$

where a or b can represent any combination of VCP from 1 to 4. Note that the following VCP geometric constraint is true for both camera and the body frame, the body frame and the camera frame VCP distances are equated as,

$$\|\mathbf{c}^w_a - \mathbf{c}^w_a\| = \|\mathbf{c}^c_a - \mathbf{c}^c_a\| \qquad [28]$$

substituting Eq. 26 into Eq. 27, the following can be derived.

$$\mathbf{L}\tilde{\boldsymbol{\beta}} = \boldsymbol{\rho} \qquad [29]$$

where $\tilde{\boldsymbol{\beta}}$ is an array containing $\tilde{\beta}_s = \beta_{nm}$, $s = 1\ldots10$ and $nm = 11, 22, 21, 33, 32, 31, 44, 43, 42, 41$ respectively. Equation 29 is the linearization form of the β polynomial, the double indices indicate

$$\tilde{\beta}_{mn} = \beta_m \beta_n \qquad [30]$$

For each of the four cases of l, the general Eq. 29 can be simplified. The simplest and most useful case is when $l = 1$ where β is a single value and can be determined as follows,

$$\begin{aligned} \beta &= \frac{\sum\limits_{a,b} \|\mathbf{v}_{1a} - \mathbf{v}_{1b}\| \|\mathbf{c}^w_a - \mathbf{c}^w_b\|}{\sum\limits_{a,b} \|\mathbf{v}_{1a} - \mathbf{v}_{1b}\|^2} \\ &= \sqrt{\frac{\mathbf{L}^T\boldsymbol{\rho}}{\mathbf{L}^T\mathbf{L}}} \end{aligned} \qquad [31]$$

For higher l cases, let,

$$\mathbf{dv}_{abnm} = (\mathbf{v}_{na} - \mathbf{v}_{nb}) \cdot (\mathbf{v}_{ma} - \mathbf{v}_{mb}) \qquad [32]$$

The matrix $\mathbf{L}$ can be generated by $l_{rs} = \mathbf{dv}_{abnm}$ entries. The r rows are combinations of the VCP distance constraint, where $r = 1\ldots6$ are $ab = 12, 13, 14, 23, 24, 34$ respectively. The s columns are the combinations of singular values, where $s = 1\ldots10$ are the nm combinations previously defined. Once the $\mathbf{L}$ matrix is formed, its SVD can be determined as,

$$\mathbf{U}_l \boldsymbol{\Sigma}_l \mathbf{V}_l^T = \mathbf{L} \qquad [33]$$

Equation 29 can be solved in the least squares for $l = 2$ and $l = 3$ respectively as,

$$\tilde{\boldsymbol{\beta}}_{l=2} = \mathbf{V}_l \left(\boldsymbol{\Sigma}_l^T \boldsymbol{\Sigma}_l\right)^{-1} \boldsymbol{\Sigma}_l^T \mathbf{U}_l^T \boldsymbol{\rho} \qquad [34]$$

$$\tilde{\boldsymbol{\beta}}_{l=3} = \mathbf{V}_l \boldsymbol{\Sigma}_l^{-1} \mathbf{U}_l^T \boldsymbol{\rho} \qquad [35]$$

For $l = 4$, the rank needed for the $\mathbf{L}$ matrix inverse exceeds the six constraints from Eq. 28. Lepetit[53] proposes a relinearization using the properties,

$$\begin{aligned} \tilde{\beta}_{mnpq} &= \beta_m \beta_n \beta_p \beta_q \\ &= \tilde{\beta}_{m'n'p'q'} \end{aligned} \qquad [36]$$

where $[m', n', p', q']$ is any combination of $[m, n, p, q]$. The accuracy for the ePnP can be further enhanced by using Gauss-Newton (GN) iteration. The formulation for ePnP+GN is as follows. Let define the square error in each of the distance constraint be ϵ_r,

$$\epsilon_r = \|\mathbf{c}^c_a - \mathbf{c}^c_b\|^2 - \|\mathbf{c}^w_a - \mathbf{c}^w_b\|^2 \qquad [37]$$

In matrix form, Eq. 37 becomes,

$$\begin{aligned} \boldsymbol{\epsilon} &= \boldsymbol{\rho}^c - \boldsymbol{\rho}^w \\ &= \mathbf{L}\tilde{\boldsymbol{\beta}} - \boldsymbol{\rho}^w \end{aligned} \qquad [38]$$

the partial rate of change of $\boldsymbol{\epsilon}$ with respect to β_l is,

$$\begin{aligned} \frac{\partial \epsilon_r}{\partial \beta_1} &= \begin{bmatrix} 2L_{r1} & L_{r3} & L_{r6} & L_{r10} \end{bmatrix} \boldsymbol{\beta} \\ \frac{\partial \epsilon_r}{\partial \beta_2} &= \begin{bmatrix} L_{r3} & 2L_{r2} & L_{r5} & L_{r9} \end{bmatrix} \boldsymbol{\beta} \\ \frac{\partial \epsilon_r}{\partial \beta_3} &= \begin{bmatrix} L_{r6} & L_{r5} & 2L_{r4} & L_{r8} \end{bmatrix} \boldsymbol{\beta} \\ \frac{\partial \epsilon_r}{\partial \beta_4} &= \begin{bmatrix} L_{r10} & L_{r9} & L_{r8} & 2L_{r7} \end{bmatrix} \boldsymbol{\beta} \end{aligned} \qquad [39]$$

where $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{bmatrix}^T$. Collecting ${}^{\partial \epsilon_r}/_{\partial \beta_l}$, where $\mathbf{g}_r$ is the individual rows of gradient matrix $\mathbf{G}$, $\mathbf{g}_r = \begin{bmatrix} {}^{\partial \epsilon_r}/_{\partial \beta_1} & {}^{\partial \epsilon_r}/_{\partial \beta_2} & {}^{\partial \epsilon_r}/_{\partial \beta_3} & {}^{\partial \epsilon_r}/_{\partial \beta_4} \end{bmatrix}$ The GN iteration is then performed as follows,

$$\beta_{z+1} = \beta_z - \left(\mathbf{G}^T\mathbf{G}\right)^{-1} \mathbf{G}^T \boldsymbol{\epsilon} \qquad [40]$$

For real-time (RT) operations, the number of iteration steps is cut off at 5 cycles, which in most instances is sufficient for convergence. Once β_l is found, it is possible to extract the body frame pose using the least square method provided by Challis[69] as follows. First find the centroid of the CP from the camera frame

$$\bar{\mathbf{P}}^c = \frac{1}{M} \sum_{k=1}^{M} \mathbf{P}^c_k \qquad [41]$$

the distance of the CP to the centroid is $\mathbf{d}^c_k = \mathbf{P}^c_k - \bar{\mathbf{P}}^c$. Recall the body frame CP is related to the camera frame CP through the extrinsic pose,

$$\mathbf{P}^c_k = \mathbf{t} + \mathbf{R}\mathbf{P}^w_k \qquad [42]$$

Substituting the centroid to CP distance for the body and camera CP into Eq. 42 generates a *cross-dispersion* matrix $\mathbf{C}$. Solving for the SVD of the cross-dispersion matrix produces the following,

$$\begin{aligned} \mathbf{C} &= \frac{1}{M} \sum_{k=1}^{M} \mathbf{d}^c_k \mathbf{d}^{wT}_k \\ &= \mathbf{U}_c \boldsymbol{\Sigma}_c \mathbf{V}_c^T \end{aligned} \qquad [43]$$

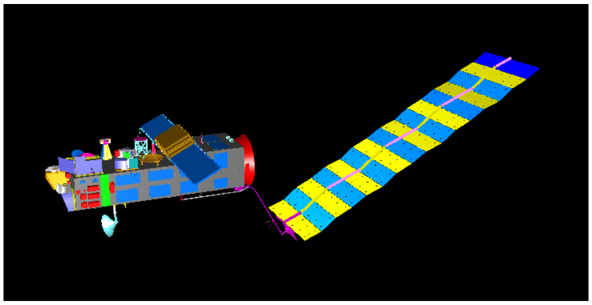


Fig. 11: Envisat 3D Model.

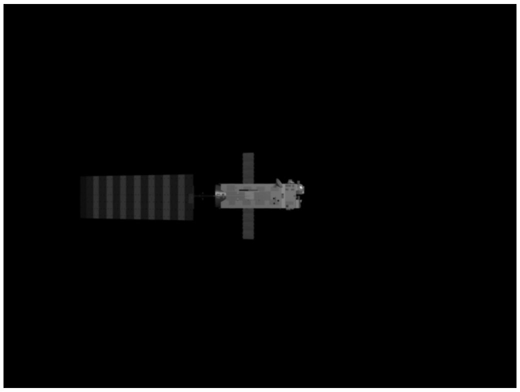


Fig. 12: Training image.

The rotation from the body frame to the camera frame can be then determined as,

$$\mathbf{R} = \mathbf{U}_c \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & |\mathbf{U}_c\mathbf{V}_c^T| \end{bmatrix} \mathbf{V}_c^T \qquad [44]$$

The translation from the camera frame to the body frame expressed in the camera frame is,

$$\mathbf{t} = \bar{\mathbf{P}}^c - \mathbf{R}\bar{\mathbf{P}}^w \qquad [45]$$

III. SIMULATION SETUP

Spacecraft pose estimation was performed using 3D synthetic images as well as using IR camera images of a satellite model. The synthetic image was taken using a virtual camera from 3D Studio Max. The target model is the Envisat as shown in Fig. 11. The 3D model under goes 360 deg yaw motion in 300 seconds, this is roughly half of the actual Envisat tumbling rate.[70] The initial zero position taken for the training image is shown in Fig. 12. The real world image is taken using an IR camera as shown in Fig. 13. The IR camera specification is provided in Tbl. 1. A model of the Radarsat shown in Fig. 14 was used for the real world image tests.

Description	Specification	Units
Resolution	320×240	pixels
Focal Length	7.5	mm
FOV	40	degrees
Thrmal Sensitivity	≤ 30	mK
Accuracy	1	deg or %
Sensor	17 UFPA VOx	micron
Spectral Response	7 – 14	micron

Table 1: ICI IR Camera Specification.

Fig. 13: ICI IR Camera.

IV. RESULTS AND DISCUSSION

Feature extraction was performed on the Envisat model using SIFT and BRIEF methods with results shown in Fig. 15. k-NN matches checked by ratio are compared with true matches that are evaluated manually. The BRIEF feature has zero real matches after 40 degrees yaw angle although k-NN matching falsely believe 20 matches still exists. SIFT match also drops to 20 as yaw angle increased to 135 degrees. However, SIFT matches by ratio are true matches showing reliability in the SIFT method. It is worth while to note prior to 15 degrees, the SIFT and BRIEF match performance are comparable. This suggest the BRIEF feature is a useful tool for frame-by-frame image matching. The match quality is defined by percentage good matches over found matches. This is the best indication of the amount of outlier that may exist using k-NN matching. Fig. 16 shows as the yaw angle increase, the BRIEF percentage good match to found match drops to zero. It also show for low rotation angles, the match quality remains relatively consistent until match ratio increase above 0.6. An image view of the keypoints and feature matches for both SIFT and Shi-Tomasi Corner/BRIEF methods are provided in Fig. 17. The figure show for small rotation angles, the keypoints and descriptors are comparable with each other. The pose estimation

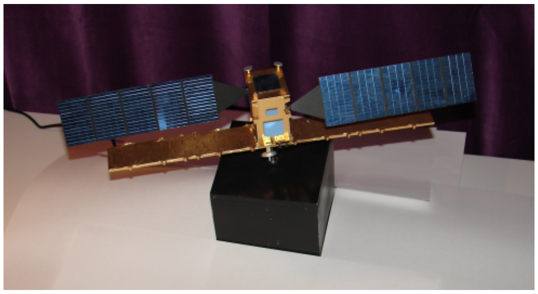


Fig. 14: Radarsat Model.

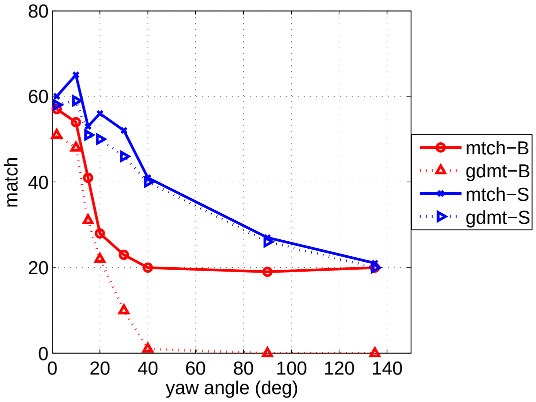


Fig. 15: SIFT vs. BRIEF matches.

of the Envisat model is shown in Fig. 18 as the spacecraft undergo 360 degrees yaw rotation. The green outline shows the initial homography mapping while the red outline show the final PnP pose estimation. The performance for computation time and pose error are shown in Fig. 19. The performance figure shows computation time, translation and rotational error. For small rotation angles, the pose error for both ePnP and *SoftPOSIT* is comparable although the computation timing using *SoftPOSIT* is by order of magnitude higher than the ePnP method. As initial misalignment becomes large, the *SoftPOSIT* method no longer provide good solution as the maximum limit of the allowed iteration is reached and the program returns without solution. For more accurate *SoftPOSIT* solution, additional RANSAC trials shall be added. Given the time required for a single *SoftPOSIT* iteration, additional RANSAC trials will by far exceed the non-iterative ePnP timing performance. By comparison, the non-iterative ePnP appraoch is superior to the iterative approach, the only draw back is it requires more than 6 CP. It is shown in Fig. 20, 20 ePnP CP is optimal for resilience to image noise. This is not a large issue since the the number of CP can be selectable from the internal 3D model. Finally, the proposed method is demonstrated using real world IR image as taken by the ICI IR camera. Figure 21 show the use of SoftPOSIT to match the RADARSAT model given an initial orientation of orientation of nearly 123 degrees.

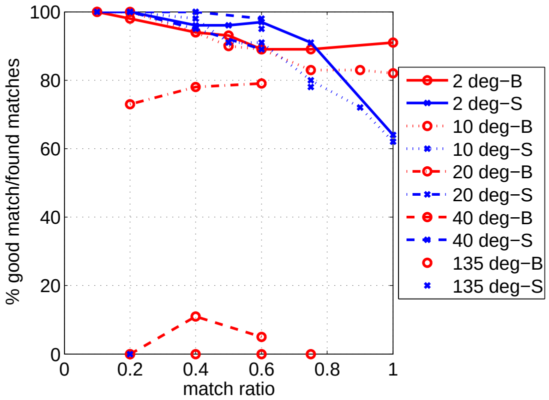


Fig. 16: Match quality relative to match ratio.

V. CONCLUSIONS

In conclusion, a method for spacecraft pose estimation was presented first by using image features to generate homography then by image projection and solving the pnp problem. This study introduced a method for image processing using histogram equalization by region (HER), compared SIFT and BRIEF features, and reviewed enhancements to initialization and limiting local minimum trapping for the *softPOSIT* algorithm. It was found twenty correspondence points from the internal model is optimal for resilient ePnP estimation to image noise. It is shown by comparison, the non-iterative PnP method is superior in timing and accuracy to iterative method for large initial condition misalignments. Future work include robust handling of cases that does not have homography solution, and improving the sentinel checker to accommodate better estimates of the target if image data is corrupt.

ACKNOWLEDGEMENTS

This research was jointly funded by the Natural Sciences and Research Council of Canada's Alexander Graham Bell Canada Graduate Scholarship CGSD3-453738-2014, Canadian Space Agency Space Technology Development Program and Ontario Centres for Excellence Voucher for Innovation and Productivity II Award 24053.

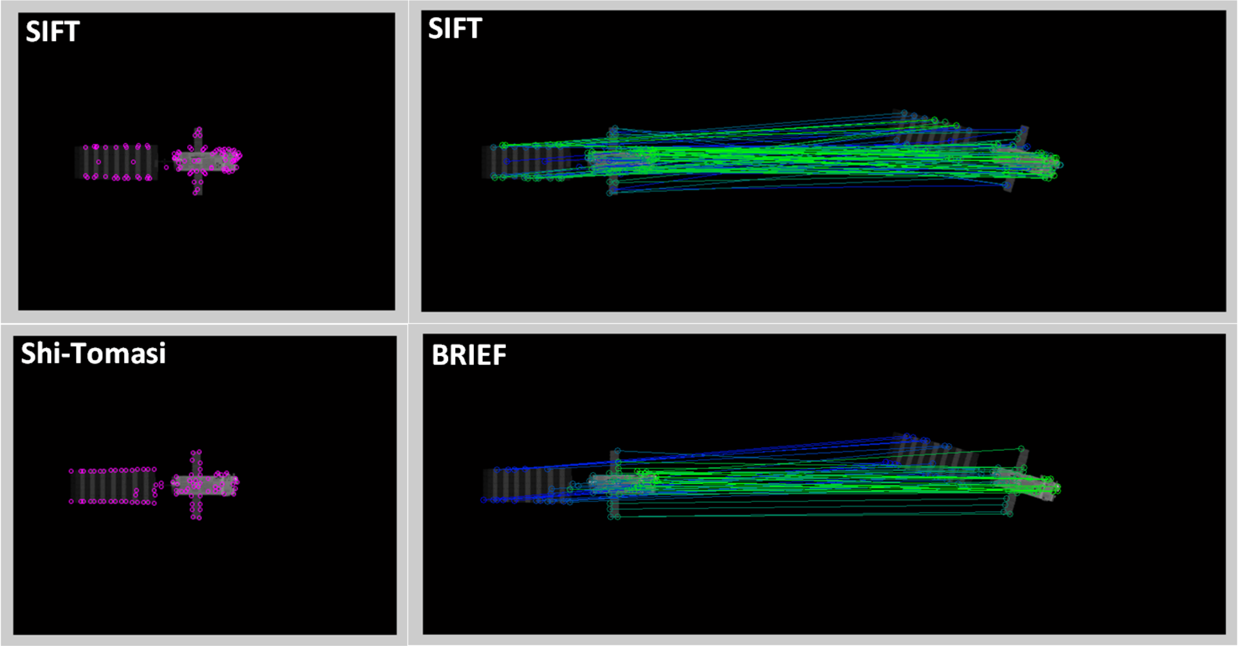


Fig. 17: Feature matching.(Top:SIFT,Bottom:BRIEF)

REFERENCES

[1] F. Khan, “Mobile internet from the heavens,” *arXiv:1508.02383*, 2015.

[2] E. Sotto and J. Bastante, “System and gnc concept for rendezvous into elliptical orbit for mars sample return mission,” in *AIAA Guidance, Navigation and Control Conference and Exhibit*, Hilton Head, South Carolina, Aug. 2007.

[3] L. Castellani, S. Llorente, M. Fernandez, J.M. Ruiz, A. Mestreau-Garreau, A. Cropp, and Santovincenzo, “Proba-3 mission,” *Space Science and Engineering, Intl. Journal of*, vol. 1, no. 4, pp. 349–366, 2013.

[4] D. Fourie, B. Tweddle, S. Ulrich, and S.-O. A., “Flight results of vision-based navigation for autonomous spacecraft inspection of unknown objects,” *Spacecraft and Rockets, AIAA Journal of*, vol. 51, no. 6, pp. 2016–2026, Feb. 2014.

[5] C. Miravet, L. Pascual, E. Krouch, and J. delCura, “An image-based sensor system for autonomous rendez-vous with uncooperative satellites,” in *Guidance, Navigation and Control Systems, 7th Intl. ESA Conf on*, Tralee, Ireland, Jun. 2008.

[6] C. Chullen and E. Blome, “H-ii transfer vehicle and the operations concept for extravehicular activity hardware,” in *Environmental Systems (ICES), 41st Intl. Conf. on*, Portland, Oregon, 2011.

[7] SpaceX, “Dragon lab,” http://www.spacex.com/ sites/ spacex/ files/ pdf/ DragonLabFactSheet.pdf Accessed: Sept 6, 2016.

[8] ESA, “Europe’s automated ship docks to the iss,” *European Space Agency News*, 2008.

[9] NASA, “Russian progress spacecraft,” http://www.nasa.gov/ mission˙pages/ station/ structure/ elements/ progress.html Accessed: Sept 6, 2016.

[10] USAF, “Fact sheet: Automated navigation and guidance experiment for local space (angels),” Jul. 2014, http://www.kirtland.af.mil/ shared/ media/ document/ AFD-131204-039.pdf Accessed:March 6, 2016.

[11] I. Mills, “Svs briefing,” in *NASA SSP DX22 Presentation*, Houston, Tx, Mar. 1999.

[12] E. S. Agency, “Esa-hso-cou-031 rev 2.0 fact sheet: H-ii transfer vehicle (htv),” http://wsn.spaceflight.esa.int/ docs/ Factsheets/ 31Accessed:Sept 6, 2016.

[13] Jena-Optronik, “Rendezvous- and docking sensor rvs,” http://www.jena-optronik.de/ de/ lageregelungssensoren/ rendezvous-und-docking-sensor-rvs.html ?file=tl˙files/ pdf/ Data

[14] R. Howard, T. Bryan, J. Lee, and B. Robertson, “Next generation advanced video guidance sensor development and test,” in *Guidance and Control Conference, 32nd Annual AAS*, Breckenridge, COa, Jan. 2009, aAS 09-064.

[15] M. R. Leinz, C. T. Chen, P. Scott, W. B. Gaumer, P. W. Sabasteanski, and M. Beaven, “Modeling, simulation, testing, and verification of the orbital express autonomous rendezvous and capture sensor system (arcss),” in *Sensors and Systems for Space Applications II, Proc. of SPIE*, vol. 6958, Bellingham WA, 2008, 6958-0C.

[16] S. Ruel, T. Luu, and A. Berube, “Space shuttle testing of the tridar 3d rendezvous and docking sensor,” *Field Robotics, Journal of*, vol. 29, no. 4, pp. 535–553, 2012.

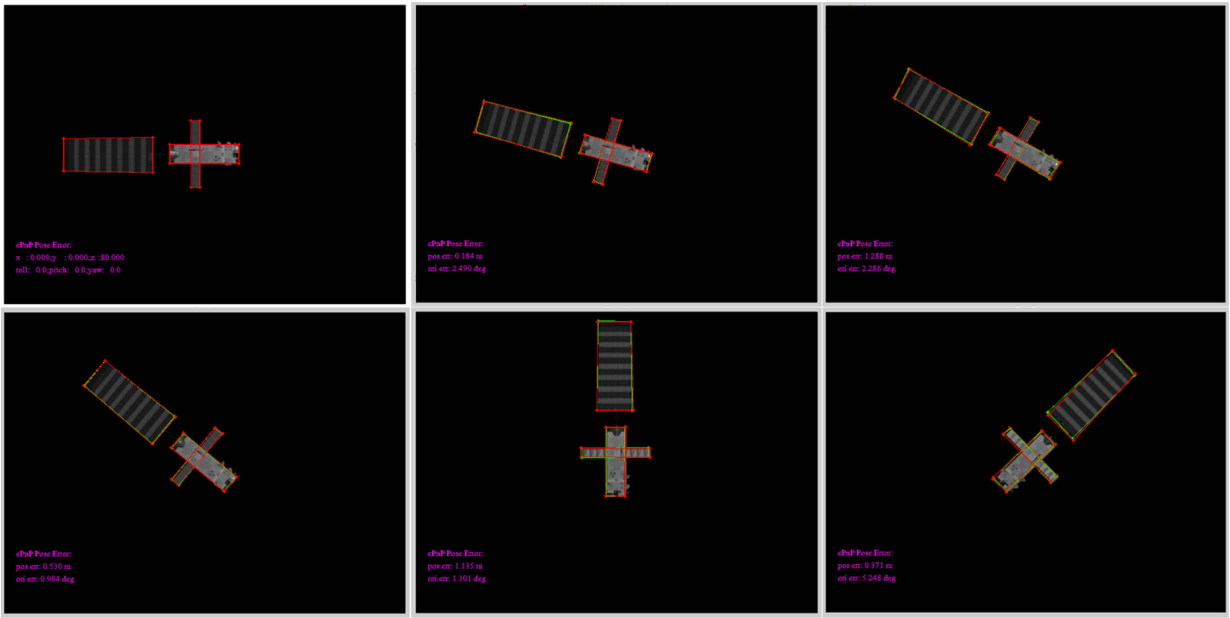


Fig. 18: Various estimated Envisat pose.(Min:0 deg,Max:135 deg)

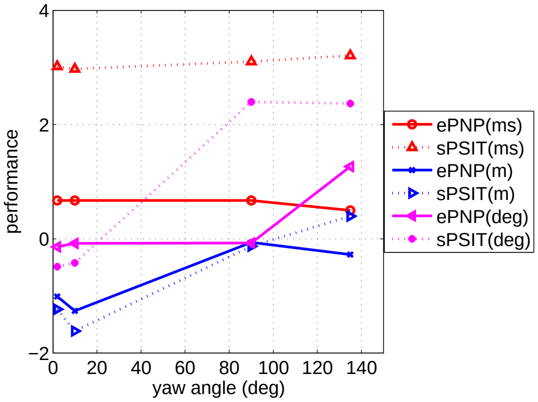


Fig. 19: $e\mathrm{P}n\mathrm{P}$ vs. SoftPOSIT performances.

[17] N. Creamer, R. Hartley, J. Obermark, C. Henshaw, S. Roderick, and J. Hope, "Autonomous release of a snagged solar array: technologies and laboratory demonstrations," *Spacecraft and Rockets, AIAA Journal of*, vol. 51, no. 1, pp. 86–95, 2014.

[18] L. Morency and R. Gupta, "Robust real-time egomotion from stereo images," in *Image Processing, Proc. Intl. Conf. on*, 2003.

[19] D. Sinclair, A. Blake, and D. Murray, "Robust estimation of egomotion from normal flow," *Computer Vision. Intl. Journal of*, vol. 13, no. 1, pp. 57–69, 1994.

[20] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *Computer Vision (IJCV), Intl. Journal of*, vol. 9, no. 2, pp. 137–154, 1992.

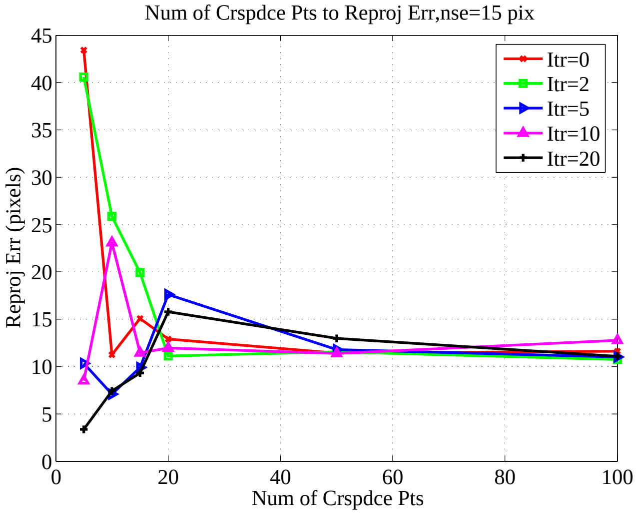
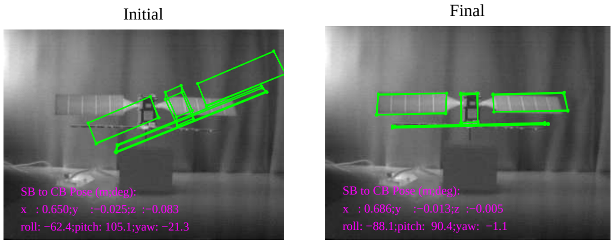


Fig. 20: Correspondence points vs. image noise.

[21] S. Augenstein and S. Rock, "Improved frame-to-frame pose tracking during vision-only slam/sfm with a tumbling target," in *Robotics and Automation, IEEE Intl. Conf on*, Shanghai, China, May 2011.

[22] D. Gennery, "Visual tracking of known three-dimensional objects," *Computer Vision (IJCV), Intl. Journal of*, vol. 7, no. 3, pp. 243–270, 1992.

[23] F. Jurie and M. Dhome, "Real time robust template match-



(a) pre-match (b) post-match

Fig. 21: *SoftPOSIT* pose estimation of a Radarsat model captured by an monocular infrared camera.

ing," in *British Machine Vision Conference (BMVC)*, 2002, pp. 123–131.

[24] B. Isard, "Condensation: Conditional density propagation for visual tracking," *Computer Vision (IJCV), Intl. Journal of*, vol. 29, no. 1, pp. –, 1998.

[25] X. Lan and D. Huttenlocher, "A unified spatio-temporal articulated model for tracking," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conf. on*, vol. 1, no. 1, 2004, pp. 722–729.

[26] P. David, D. Dementhon, R. Duraiswami, and H. Samet, "Softposit: simultaneous pose and correspondence determination," *Computer Vision, Intl. Journal of*, vol. 59, no. 3, pp. 259–289, 2004.

[27] J. Kelsey, J. Byrne, M. Cosgrove, S. Seereeram, and R. Mehra, "Vision-based relative pose estimation for autonomous rendezvous and docking," in *Aerospace Conference, IEEE*, 2002.

[28] J. Diaz and M. Abderrahim, "Visual inspection system for autonomous robotic on-orbit satellite servicing," in *Advanced Space Technologies for Robotics and Automation (ASTRA), Proc. of 9th ESA Workshop on*, Noordwijk, Netherlands, Nov. 2006.

[29] G. Arantes Jr., "Rendezvous with a non-cooperating target," Master's thesis, The Dept. of Production Engineering, Univ. of Bremen, Bremen, Germany, 2011.

[30] S. Sharma and S. D'Amico, "Comparative assessment of techniques for initial pose estimation using monocular vision," *Acta Astronautica*, vol. 123, pp. 435–445, 2016.

[31] J. Chen, X. Luo, H. Liu, and F. Sun, "Cognitively inspired 6d motion estimation of a noncooperative target using monocular rgb-d images," *Cognitive Computation*, vol. 8, pp. 1–9, 2016.

[32] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Computer Vision, Intl. Journal of*, vol. 60, no. 2, pp. 91–110, 2004.

[33] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Graphics and Image Processing*, vol. 24, no. 6, pp. 381–395, 1981.

[34] S. Pizer, R. Johnston, J. Ericksen, B. Yankaskas, and K. Muller, "Contrast-limited adaptive histogram equalization speed and effectiveness," in *Visualization in Biomedical Computing, Proc. of the 1st Conf. on*, 1990, pp. 337–345.

[35] J. Lucas, B. Calef, and K. Knox, "Image enhancement for astronomical scenes," in *SPIE Optical Engineering and Applications*, Sep. 2013, p. 885603.

[36] H. Bay, A. Ess, T. Tuytelaars, and L. Gool, "Speeded-up robust features (surf)," vol. 110, no. 3, pp. 346–359, Jun. 2008.

[37] P. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *British Machine Vision Conference (BMVC)*, Bristol, UK, 2013.

[38] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: binary robust independent elementary features," in *Computer Vision (ECCV), 11th European Conference on*, Heraklion, Crete, 2010.

[39] J. Shi and C. Tomasi, "Good features to track," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conf. on*, Seattle, WA, Jun. 1994.

[40] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: an efficient alternative to sift or surf," in *Computer Vision (ICCV), IEEE 15th Intl Conf. on*, Barcelona, Spain, Nov. 2011.

[41] R. Duda, *Pattern Classification*, 2nd ed. John Wiley and Son, 2012.

[42] E. Trucco and A. Verri, *Introductory Techniques for 3D Computer Vision*. Prentice-Hall, Inc, 1998.

[43] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[44] C. Lu, G. Hager, and E. Mjolsness, "Fast and globally convergent pose estimation from video images," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 22, no. 6, pp. 610–622, 2000.

[45] G. Schweighofer and A. Pinz, "Globally optimal o(n) solution to the pnp problem for general camera models." in *British Machine Vision Conference (BMVC)*, Leeds, UK, 2008.

[46] B. Fan, Y. Du, and Y. Cong, "Robust and accurate online pose estimation algorithm via efficient 3d collinarity model," *Computer Vision, The Institution of Engineering and Technology (IET)*, vol. 7, no. 5, pp. 382–393, 2013.

[47] H. Zhou, T. Zhang, and W. Lu, "Vision-based pose estimation from points with unknown correspondences," *Image Processing, IEEE Trans. on*, vol. 23, no. 8, pp. 3468–3478, 2014.

[48] P. Sun, C. Sun, Q. Li, W, and W. P., "A new pose estimation algorithm using a perspective-ray-based scaled orthographic projection with iteration," *PLOS ONE*, vol. 10, no. 7, p. e0134029, 2015.

[49] S. Urban, J. Leitloff, and S. Hinz, “Mlpnp-a real-time maximum likelihood solution to the perspective-n-point problem,” *arXiv:1607.08112*, 2016.

[50] X. Gao, X. Hou, J. Tang, and H. Cheng, “Complete solution classification for the perspective-three-point problem,” *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 25, no. 8, pp. 930–943, 2003.

[51] L. Kneip, D. Scaramuzza, and R. Siegwart, “A novel parameterization of the perspective-three-point problem for a direct computation of absolute camera position and orientation,” in *Computer Vision and Pattern Recognition (CVPR), IEEE Conf. on*, Colorado Springs, CO, Jun. 2011.

[52] F. Moreno-Noguer, V. Lepetit, and P. Fua, “Accurate non-iterative o(n) solution to the pnp problem,” in *Computer Vision (ICCV). IEEE 11th Intl. Conf. on*. IEEE, 2007, pp. 1–8.

[53] V. Lepetit, “Epnp: an accurate o(n) solution to the pnp problem,” *Computer Vision, Intl. Journal of*, vol. 81, no. 2, pp. 155–166, 2009.

[54] J. Gao and Y. Zhang, “An improved iterative solution to the pnp problem,” in *Virtual Reality and Visualization (ICVRV), Intl. Conf. on*, 2013.

[55] L. Ferraz, X. Binefa, and F. Moreno-Noguer, “Very fast solution to the pnp problem with algebraic outlier rejection,” in *Computer Vision and Pattern Recognition (CVPR), Proc. of the IEEE Conf. on*, 2014, pp. 501–508.

[56] S. Li, C. Xu, and M. Xie, “A robust o(n) solution to the perspective-n-point problem,” *Pattern Analysis Machine Intelligence (PAMI), IEEE Trans. on*, vol. 34, no. 7, pp. 1444–1450, 2012.

[57] Y. Zheng, S. Sugimoto, and M. Okutomi, “Aspnp: An accurate and scalable solution to the perspective-n-point problem,” *Information and Systems, Trans. on*, vol. 96, no. 7, pp. 1525–1535, 2013.

[58] Y. Zheng, Y. Kuang, S. Sugimoto, K. Astrőm, and M. Okutomi, “Revisiting the pnp problem: A fast, general and optimal solution,” in *Computer Vision (ICCV). IEEE 11th Intl. Conf. on*. IEEE, 2013, pp. 4321–4328.

[59] L. Kneip and P. Furgale, “Upnp: An optimal o(n) solution to the absolute pose problem with universal appilcability,” in *Computer Vision (ECCV), 13th European Conf. on*. IEEE, 2014, pp. 127–142.

[60] G. Nakano, “Globally optimal dls method for pnp problem with cayley parameterization.”

[61] J. Hesch and S. Roumeliotis, “A direct least-squares (dls) method for pnp,” in *Computer Vision (ICCV), IEEE 15th Intl Conf. on*, Barcelona, Spain, Nov. 2011.

[62] J. Shi and S. Ulrich, “Softposit enhancements for monocular camera spacecraft pose estimation,” in *Methods and Models in Automation and Robotics (MMAR), IEEE Intl. Conf. on*, Międzyzdroje, Poland, Aug. 2016.

[63] S. Kirkpatrick, C. Gelatt, and M. Vecchi, “Optimization by simulated annealing,” *Science, New Series*, vol. 220, no. 4598, pp. 671–680, 1983.

[64] D. Dementhon and L. Davis, “Model-based object pose in 25 lines of code,” *Computer Vision, Intl. Journal of*, vol. 15, no. 1, pp. 123–141, 1995.

[65] S. Gold, A. Rangarajan, C. P. Lu, S. Pappu, and E. Mjolsness, “New algorithms for 2d and 3d point matching: pose estimation and correspondence,” *Pattern Recognition*, vol. 31, no. 8, pp. 1019–1031, 1998.

[66] R. Sinkhorn, “A relationship between arbitrary positive matrices and doubly stochastic matrices,” *The Annals of Mathematical Statistics*, vol. 35, no. 2, pp. 876–879, 1964.

[67] K. Jager, M. Hebel, and K. Bers, “Automatic 3d object pose estimation in ir image sequences for forward motion application,” in *Automatic Target Recognition XIV, Proc. of SPIE*, vol. 5426, Bellingham, WA, 2004.

[68] J. Diaz and M. Abderrahim, “Modified softposit algorithm for 3d visual tracking,” in *IEEE International Symp. on Intelligent Signal Processing*, Oct. 2007.

[69] Challis, “A procedure for determining rigid body transformation parameters,” *Biomechanics, Journal of*, vol. 28, no. 6, pp. 733–737, 1995.

[70] D. Kucharski, G. Kirchner, F. Koidl, R. Fan C., Carman, C. Moore, A. Dmytrotsa, M. Ploer, G. Bianco, M. Medvedskij, A. Makeyev, G. Appleby, M. Suzuki, J. Torre, Z. Zhang, L. Grunwaldt, and F. Qu, “Attitude and spin period of space debris envisat measured by satellite laser ranging,” *Geoscience and Remote Sensing, IEEE Trans. on*, vol. 52, no. 12, pp. 7651–7657, 2014.