doi: 10.82124/CEAS-GNC-2026-082 CEAS-GNC-2026-082

AI-Based Asynchronous Sensor Data Fusion for Autonomous Relative Navigation around Asteroids

Iain Hall PhD. Candidate, The University of Strathclyde, Glasgow, United Kingdom. iain.hall@strath.ac.uk

Jinglang Feng Associate Professor, The University of Strathclyde, Glasgow, United Kingdom. jinglang.feng@strath.ac.uk

Jesús Gil Fernández GNC Engineer, European Space Agency, ESTEC, Noordwijk, Netherlands. jesus.gil.fernandez@esa.int

Massimilano Vasile Professor, The University of Strathclyde, Glasgow, United Kingdom. massimiliano.vasile@strath.ac.uk

ABSTRACT

The estimation of relative position and attitude (Pose) is necessary for enabling autonomous close proximity operations of spacecraft visiting asteroids. This can be achieved by tracking surface features of the target asteroid, but the tracking of these features is challenging due to the extreme illumination conditions and uncertainties in the target's motion and shape. To address these challenges we apply Deep Learning networks to track keypoints in visible and thermal images of an asteroid. We then develop a factor graph based Visual Odometry method to estimates the relative pose of the spacecraft using the tracked keypoints and a LIDAR range measurement to remove scale ambiguity. Sensor data fusion of visible and and thermal images is achieved using the factor graph to enable robustness to the extreme illumination conditions. We show that DL and factor graph based VO are able to accurately estimate the relative pose at low sun phase angles, while thermal images and sensor fusion provide robustness to higher sun phase angles. Finally we test the developed method on real images of Ryugu from the spacecraft Hayabusa2, and successfully demonstrate the the DL networks can adapt to real images.

Keywords: Artificial Intelligence, Deep Learning, Sensor Data Fusion, Asteroid Navigation

Nomenclature

Ω	=	Angular velocity of asteroid
$\mathcal{F}_I$,	=	Inertial reference frame
$\mathcal{F}_{sc}$,	=	Spacecraft/camera reference frame
$\mathcal{F}_A$,	=	Asteroid fixed reference frame
R_{sc}	=	Rotation of the spacecraft/camera fixed frame
R_A	=	Rotation of the asteroid fixed frame
r_{sc}	=	Spacecraft position in inertial reference frame
v_{sc}	=	Spacecraft velocity in inertial reference frame

L = Position of Landmark in asteroid fixed frame
$D_{asteroid}$ = Asteroid mean diameter
X = Spacecraft asteroid state
K = Camera intrinsic matrix

1 Introduction

Missions to explore asteroids are complex and require advanced navigation systems to be developed for their success. The navigation system, which is responsible for estimating the relative state of the spacecraft to the target asteroid has several requirements. The navigation system must be autonomous due to the long round trip communication time with Earth. It must be able to estimate the relative position and attitude of the asteroid. It must be robust to the extreme lighting conditions and uncertainties in the asteroid's motion and shape.

Previous asteroid exploration missions have developed unique asteroid relative navigation systems, using vision based navigation techniques. JAXA's Hayabusa2 mission used fiducial markers that were deployed to the surface of the asteroid Ryugu before descent operations began[1]. NASA's OSIRIS-REx mission developed a natural feature tracking system that used a shape model of the target asteroid, Bennu, to render an image from the current estimate spacecraft pose. The difference between the rendered image and an image captured at the same time is then used to update the estimation of pose[2]. Hera, ESA's mission to visit the Didymos binary asteroid system will employ Lambertian sphere correlation[3] for relative position estimation and Kanade–Lucas–Tomasi (KLT) [4] feature tracking for pose estimation [5].

All of these navigation systems rely on conventional computer vision techniques, however AI based methods, that use Deep Learning (DL) significantly outperform conventional methods for many computer vision problems including image segmentation and classification[6], and feature extraction and matching[7]. DL networks that have been developed for terrestrial feature tracking have been demonstrated to be effective at extracting and matching features from images of asteroids [8, 9]. Different Convolutional Neural Network (CNN) Architectures are used to extract and describe image features from asteroid images and matched using mutual nearest neighbour matching. The features extracted by the CNN architectures are demonstrated to outperform conventional feature extractors (ORB[10] and SIFT[11]) for precision and recall.

Image feature extraction and matching alone are not sufficient for pose estimation, they must be used as part of Visual Odometry (VO) and visual Simultaneous Localization And Mapping (vSLAM) methods to track a camera's pose relative to an environment. A two step process is used, consisting of a front end, responsible for matching image features, and a back end, responsible for mapping and estimating relative motion from the matched features[12]. VO and vSLAM methods have been developed for asteroid navigation and two main types of methods have been explored, filter based methods and graph based methods. Razgus et al. use a dual quaternion quantification of the state with and Extended Kalman Filter (EKF)[13]. Ma et al. use an EKF to develop a hovering vision based control around an asteroid[14]. A factor graph based vSLAM method has been developed by Dor et al. [15], using ORB features to estimate relative pose of a spacecraft.

The feature extraction methods explored in [8, 9] only use visible images, but more recent missions to asteroids, including Hayabusa2, Hera, and RAMSES have thermal cameras. This allows the potential use of sensor data fusion, the process by which data from multiple different sensor sources is combined to provide an improved measurement[16]. We explore how thermal and visible images can be fused to improve the relative pose estimation and its effect on the navigation's robustness to extreme lighting conditions. Visible images provide higher quality images with clear shadowing, while thermal images

provided robustness to poor illumination conditions. An asynchronous image fusion method is considered because it is not possible to guarantee the simultaneous capture of images onboard

In this work we explore how DL networks can be used to extract and match keypoints in visible and thermal images. We then develop a factor graph implementation to estimate relative pose from these matched features and to carry out sensor data fusion. With the trained DL networks and factor graph we have developed both the front end and back end of a VO method for navigation around an asteroid and we test it on images of the asteroid Ryugu.

2 Method

The method covers the DL networks employed, the factor graph implementation, the sensor data fusion architecture, and how they are combined for VO to estimate the relative pose of the spacecraft to the asteroid. Figure 1 shows the basic VO pipeline in which keypoint matches are used to add measurement factors to asteroid and spacecraft states, while propagation factors are used to add the relative dynamics between states.

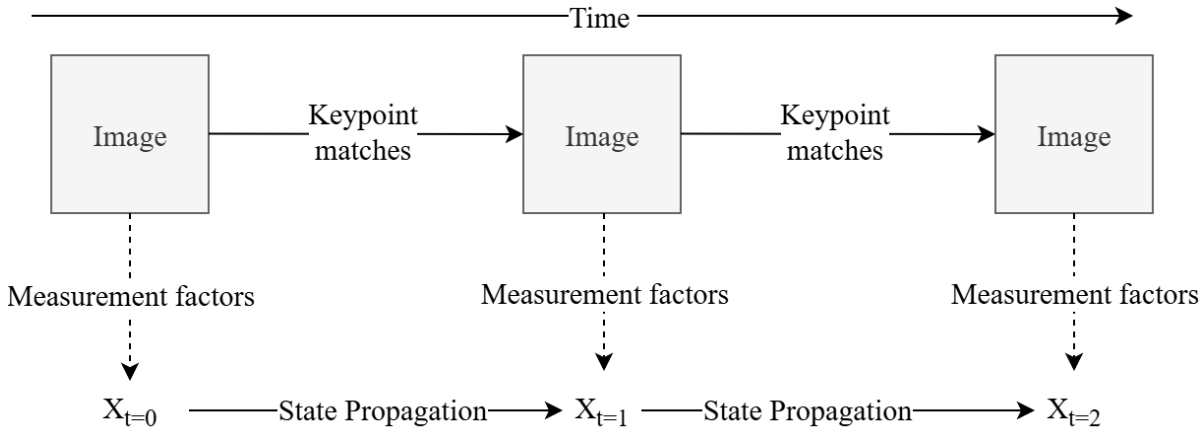

Fig. 1 VO pipeline using matched keypoints from the DL network to update state, X.

2.1 Deep Learning Extraction and Matching of Keypoints

The DL networks are responsible for tracking landmarks on the surface of the asteroid. Landmarks are fixed surface features on the asteroid which are tracked using the matching of keypoints, which represent the location of the landmark in a given image. Two networks are used in the process of extracting and matching keypoints. The Superpoint network [17] is used to extract and describe keypoints in a single image. It is a CNN architecture that encodes the image to a feature space and then decodes it to keypoints and descriptors. Keypoints are learned salient points that are identifiable between images, and descriptors are vectors which describe the landmark a keypoint is on, so that it can be matched between images. The Lightglue network[18] is then used to match the extracted keypoints from a pair of images. It is a transformer architecture which uses self-attention on descriptors and keypoints from the same image and cross-attention for descriptors and keypoints between the images to match them. The combination of the two networks is shown in Figure 2 where keypoints and descriptors are extracted from a visible and thermal image before being matched by the Lightglue network.

To train the DL networks a synthetic dataset of images with ground truth correspondence between pixel locations was generated as follows. A set of pose pairs were generated with the characteristics defined in Table 2, while the parameters physical meanings are shown in Figure 3. Initial pose sun phase angle is the angle between the position vector at pose 1 and the sun position vector. The Off Nadir rotation is the magnitude of the rotation of the pose away from the nadir, as the result of 3 successive Euler angle

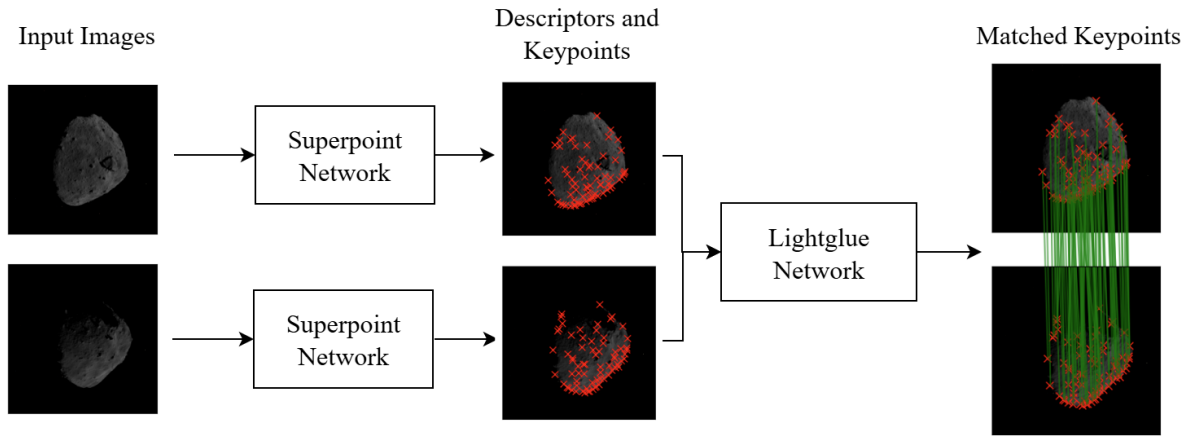

Fig. 2 Diagram of the structure used to extract and match keypoints using the Superpoint and LightGlue models

Table 2 Details of the parameters used when generating pose pairs for training the DL networks

Parameter	Minimum	Maximum
Initial Pose Range (km)	1	8
Initial Pose Sun Phase Angle (^{o})	0	50
Final Pose translation (%)	0	40
Off Nadir rotation (^{o})	0	3

rotations. The final pose translation is how large a change in position can occur between pose 1 and pose 2 as a percentage of the original position. A Synthetic dataset of 30,000 pairs of poses is generated with some random camera pose change between the pair. For each pose in the pair a thermal and visible image is generated using ESA's Planet and Asteroid Natural Scene Generation Utility (PANGU) software[1]. The ground truth correspondence of the keypoints extracted in these images can be found using the change in pose and a depth map for each image. The asteroid chosen for image generation is Ryugu due to the availability of real visible and thermal images of Ryugu from the Hayabusa2 mission from NASA's Planetary Data System[19].

Two different networks are trained using the same architecture but different training inputs. Visible only image inputs are used to train a visible to visible keypoint matcher. Thermal only image inputs are used to train a thermal to thermal keypoint matcher. The networks are trained for 30 epochs with a batch size of 32 pose pairs on the synthetic dataset with 26,000 pose pairs used for training and 4,000 reserved for validation.

2.2 Factor graph VO fundamentals

Factor graphs provide a powerful tool for solving VO and SLAM problems, by allowing the combination of many non linear measurements into a single graph, and with the ability to effectively optimise the full estimated trajectory to match these measurements. A factor graph consist of variables, stored in nodes, and factors which connect nodes. The variables are used to store the values to be estimated, such as the spacecrafts position or a landmarks location on the surface of the asteroid. The factors store measurement information and connect the variables which effect that measurement, Figure 4 shows how landmark and state nodes are connected using measurement and propagation factors. A factor will have a measurement, z, and a function, h, which estimates the measurement, $\hat{z}$, from the connected variables.

[1]STAR-Dundee: Planet and asteroid natural scene generation utility, 2025. https://pangu.software/.

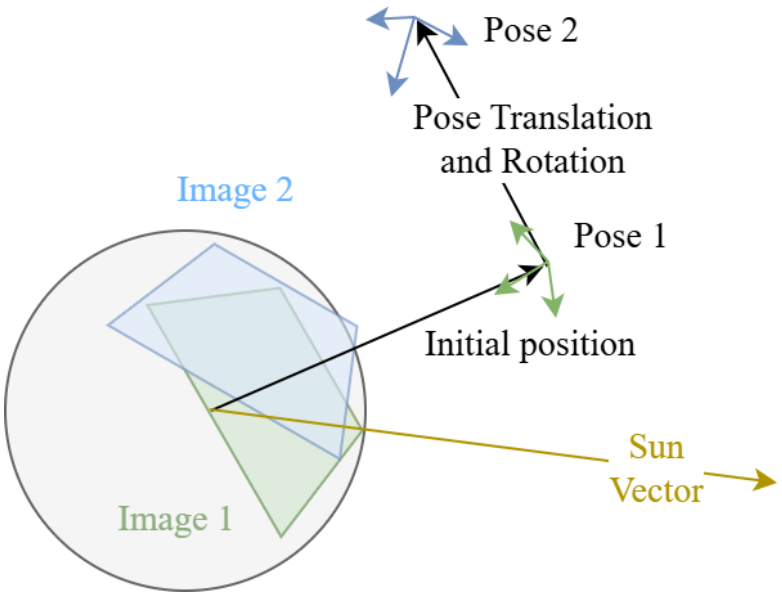

Fig. 3 Diagram showing a pair of poses with the relevant details used for the dataset generation. Sun phase angle is the angle between the initial position vector and the sun vector, Off nadir rotation is the angle between the nadir direction and the z direction of the pose.

Navigation using a factor graph is achieved by finding the state which minimises the error between z and $\hat{z}$. More specifically the total error of the graph is found using Equation 2, where Z is the full set of measurements, X is the full set of variables, and $P_{i,j}$ is the covariance of a given factor connecting variables i and j. The minimisation of $E(Z,X)$ is done using a least squares optimisation method, using the Jacobians of the measurement functions [20]. We now develop a factor graph implementation applicable to asteroid relative navigation using the matched keypoints from the DL networks.

$$e_{ij} = \hat{z}_{ij} - z_{ij} \tag{1}$$

$$E(Z,X) = \sum_{i,j,i\neq j} e_{i,j} P_{i,j}^{-1} e_{i,j}^{T} \tag{2}$$

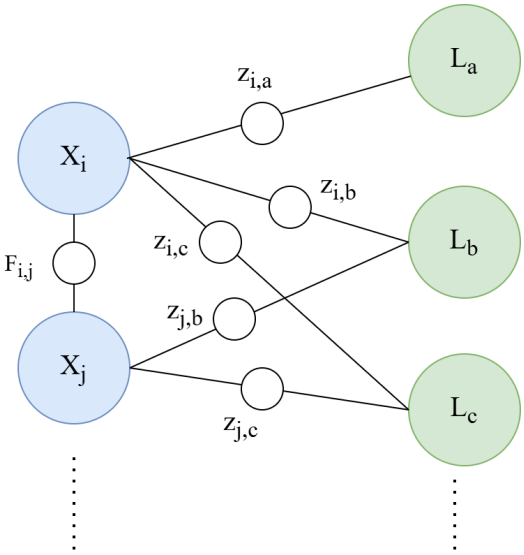

Fig. 4 Pose landmark graph, showing relation between states X_i, and landmarks L_j using measurements $z_{i,j}$, where the measurements are keypoint position measurements at different states.

2.3 Factor Graph Implementation

The values necessary for navigation of the spacecraft are the spacecrafts position and velocity in the inertial reference frame, the asteroid reference frame's attitude to the inertial frame, and the spacecraft's attitude to the inertial frame. The following simplifying assumptions are made to simplify the design of the factor graph. The first assumption is that the spacecraft's attitude to an inertial reference frame can be found to a high degree of accuracy from the attitude determination and control system (ADCS). The second assumption is that the asteroid's rotation is constant and around a single axis. A final assumption is made that the features on the surface do not move relative to an asteroid fixed rotating reference frame.

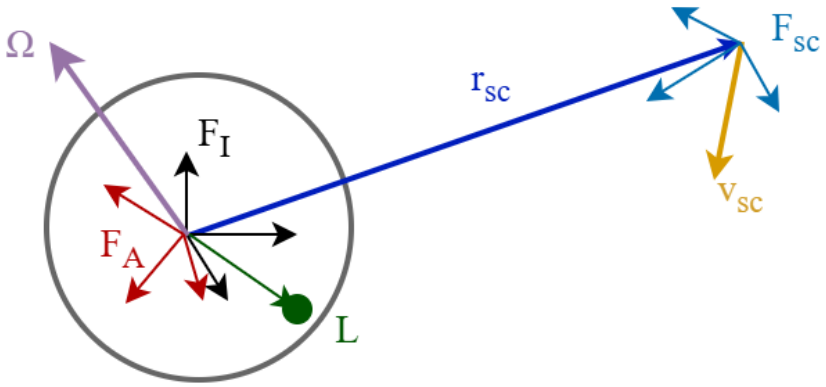

Fig. 5 Diagram of the reference frames and vectors which are defined for the estimation of relative pose.

In Figure 5 the relevant variables are shown: the asteroid reference frame, $\mathcal{F}_A$, defined relative to the inertial frame by the rotation R_A; The asteroid's angular velocity, Ω, defines the change in R_A with time; The spacecraft (camera) reference frame, $\mathcal{F}_{sc}$, defined relative to the inertial frame by the rotation R_{sc}; the relative position of the spacecraft to the centre of the asteroid in the inertial reference frame, r_{sc}; the spacecraft's velocity in the inertial frame, v_{sc}; the position of the landmark in the asteroid fixed reference frame, L. The values R_A,r_{sc}, v_{sc} form the spacecraft asteroid state, X, and are estimated for each image input, Ω also forms part of the spacecraft asteroid state—but because it is assumed to have a constant value is only estimated once, R_{sc} is assumed to be known from the ACDS system and is not estimated, and the set of landmarks L form a map. To estimate the values of these variables and therefore the relative pose it is necessary to develop a factor relating them. This is done using the matched keypoint information. A chain of matched keypoints belong to the same single landmark and for an image at a given time the expected pixel position of the keypoint corresponding to the landmark in the camera frame is given by Equation 3, where K is the camera intrinsic matrix, R_{sc} is the rotation matrix from the spacecraft (camera) frame to the inertial frame, R_A is the rotation from the asteroid frame to the inertial frame, L is the position of the landmark in the asteroid frame and r_{sc} is the position of the spacecraft relative to the asteroid in the inertial frame.

$$h(X, L) = K{R_{sc}}^{-1}(R_A L - r_{sc}) \tag{3}$$

This is sufficient for keypoint measurement factors but it is desirable to use knowledge of the spacecraft dynamics to help track the relative pose. Factors relating the state of the spacecraft and asteroid at different times are shown in Figure 6. Propagation factors, $F_{i,j}$ are used to relate the state of the spacecraft at different times, and are developed using the 2 body problem formulation. Rotation factors, $G_{i,j}$, are developed using the asteroid state at attitude at a given time and the asteroid's estimated angular velocity.

Propagation factors are implemented using the error between a spacecraft state at a node j, x_j, and the propagated spacecraft state from a prior node i to node j, $\hat{x}_j$. This is shown in Equation 4, where $\hat{x}_j$ is calculated as shown in Equation 5. The spacecraft state for a given node is defined in Equation 6, where

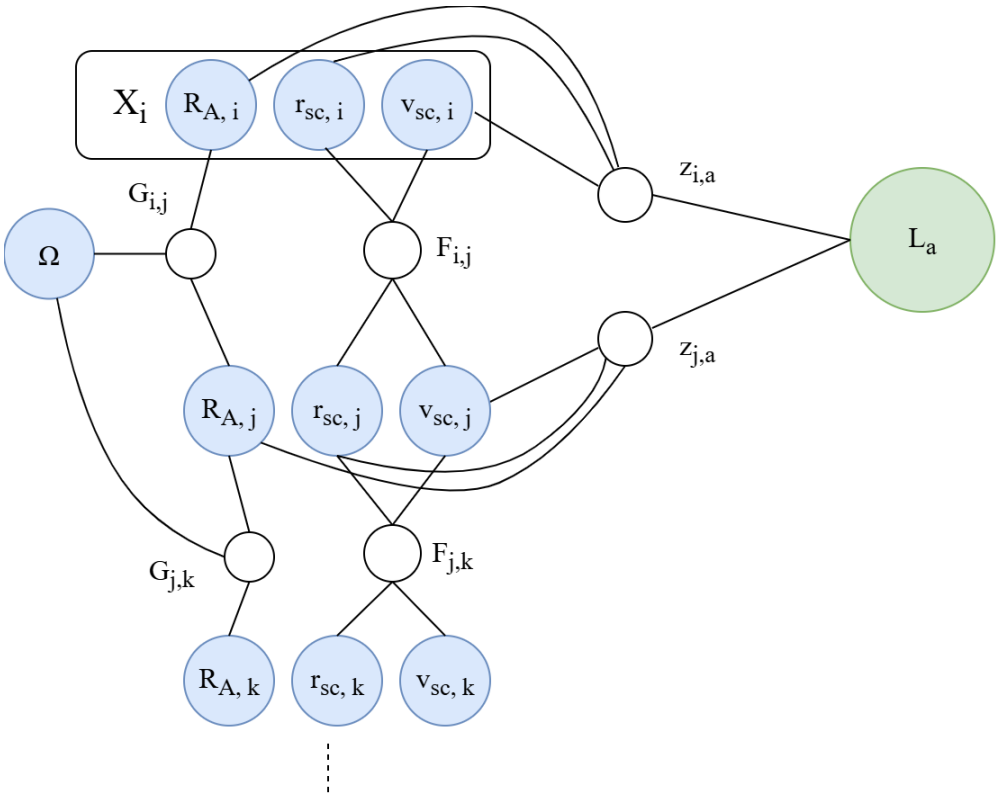

Fig. 6 Factor graph of propagation for three states, where each state X contains the spacecraft position, r_{sc}, and velocity, v_{sc}, and the asteroid's attitude R_A. The factor $G_{i,j}$ relates the rotation and angular velocity states, $F_{u,j}$ relates the spacecraft's motion in the inertial frame, $z_{i,a}$ and $z_{j,b}$ are measurement factors from a matched keypoint.

v_{sc} is the spacecraft's velocity in the inertial frame. Equation 7 shows F_{sc} is the state transition matrix according to the two body problem, where μ is the standard gravitational parameter of the asteroid and dt is the time step between nodes i and j .

$$e_{ij} = \hat{x}_j - x_j \tag{4}$$

$$\hat{x}_j = f(x_i) = F_{sc} x_i \tag{5}$$

$$x = [r_x, r_y, r_z, v_x, v_y, v_z]^T \tag{6}$$

$$F_{sc} = \begin{bmatrix} 1 & 0 & 0 & dt & 0 & 0 \\ 0 & 1 & 0 & 0 & dt & 0 \\ 0 & 0 & 1 & 0 & 0 & dt \\ -\frac{\mu}{|r|^3} dt & 0 & 0 & 1 & 0 & 0 \\ 0 & -\frac{\mu}{|r|^3} dt & 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{\mu}{|r|^3} dt & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

The asteroid rotations, R_A, describe the 3D rotation between the inertial frame and the asteroid fixed frame and belong to the group of 3D rotations, $R_A \in SO(3)$. $SO(3)$ is a lie group used to describe rotations in 3D space using a matrix with a determinant of 1 and has a lie algebra $\mathfrak{so}(3)$ which describe an equivalent rotation using a vector with a direction parallel to the axis of rotation and a magnitude equal to the size of the rotation. Relating the Lie group $SO(3)$ and lie algebra $\mathfrak{so}(3)$ are a logarithmic map, $log(Y)$, and Exponential map $exp(y)$ which exist such that:

$$Y = exp(y), \quad y = log(Y), \quad Y \in SO(3), \quad y \in \mathfrak{so}(3) \tag{8}$$

The rotation factors, G, relate two consecutive asteroid rotation frames using the estimated angular velocity, Ω, and the time step, dt. The error associated with a given rotation factor is described in Equation 9, where the error, $e_{ij} \in \mathfrak{so}(3)$, is the difference in local coordinates of $R_{A,j} \in SO(3)$ the estimated asteroid rotation matrix at j and $\hat{R}_{A,j} \in SO(3)$ is the propagated asteroid rotation matrix. Finally the $\ominus$ operator represents the general difference operator and its formulation for $SO(3)$ is described in 10.

$$e_{i,j} = \hat{R}_{A,j} \ominus R_{A,j} \tag{9}$$

$$\hat{R}_{A,j} \ominus R_{A,j} = logMap({R_{A,j}}^{-1} \hat{R}_{A,j}) \tag{10}$$

$\hat{R}_{A,j}$ is propagated from $R_{A,i}$, Ω, and dt as shown in the Equations 11 and 12, note that $\Omega dt \in \mathfrak{so}(3)$.

$$\hat{R}_{A,j} = g(R_{A,i}, \Omega) \tag{11}$$

$$g(R_{A,i}, \Omega) = expMap(\Omega dt) R_{A,i} \tag{12}$$

Two additional types of factors are included, prior factors and unary factors. Prior factors are used to add a-priori estimated state information information: the initial position, initial velocity, initial asteroid rotation, and angular velocity. Unary factors are used to add measurements which are only dependent on a single variable. The unary measurement factor is used to include a LIDAR range measurement, with the measurement function shown in Equation 13, where $D_{asteroid}$ is the mean diameter of the target asteroid

$$h_{range}(r) = ||r|| - \frac{D_{asteroid}}{2} \tag{13}$$

2.4 Sensor Data Fusion

To carry out the asynchronous sensor data fusion of visible images and thermal images the basic pipeline for developing the factor graph from matched keypoints is modified. Sensor data fusion is integrated into the the factor graph using matches between the two different image types to update a common state. Figure 7 shows how the asynchronous implementation of VO where the keypoint matches between the visible images and thermal images are found using the respective DL network. These then form two sets of landmarks which are used to update a common set of asteroid spacecraft states, X. This method allows for sensor data from visible and thermal images to be fused without any requirements on the order of the image types or the time between the different image types.

2.5 Implementation details

The VO implementation is achieved using the Georgia Tech Smoothing And Mapping (GTSAM) Library[21], with GTSAM's non-linear factor graph used to construct the graph and Levenberg–Marquardt algorithm to optimise the values. Incorrect keypoint matches will negatively effect the pose estimation, so to reduce the number of incorrect keypoint matches a minimum consecutive matches metric is used before a landmark can be included in the factor graph. Additionally a Huber loss metric [22] is used for the keypoint error loss, which improves the factor graph's robustness to outliers. The factor graph are optimised a-posteriori after all measurements have been made, which is representative of an offline pose estimation scenario.

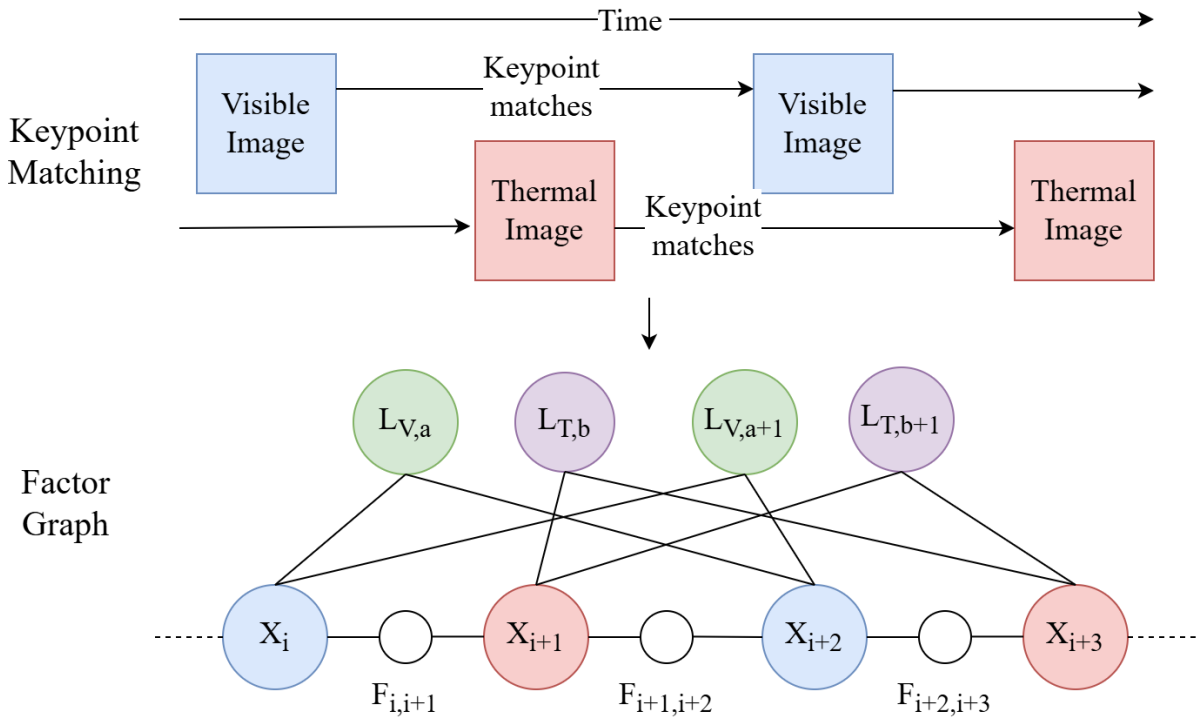

Fig. 7 VO pipeline used for asynchronous fusion of visible and thermal keypoint matches. Keypoint matches are made between individual image types forming and a factor graph is made with a set of visible landmarks, L_V, and thermal landmarks L_T, and spacecraft asteroid States, X, which are linked with propagation factors, to achieve the sensor fusion.

3 Results

3.1 DL Benchmarks

Benchmarks are used to assess the performance of the DL models individually and for the full VO method. The metrics used to assess the DL networks are Precision, Recall, and the Average Precision. The Precision, P, is defined in equation 14, where correct matches are the true keypoint matches that are identified by the network and incorrect matches are the keypoint matches which are identified but are between different keypoints. The Recall, R, is defined in equation 15, where missed matches are when a keypoint exists in both images but is not matched by the DL network. The average precision, AP, is a combined metric that measures the area under the precision recall curve and is defined in 16.

$$P = \frac{\text{Correct Matches}}{\text{Correct Matches + Incorrect Matches}} \tag{14}$$

$$R = \frac{\text{Correct Matches}}{\text{Correct Matches + Missed Matches}} \tag{15}$$

$$AP = \int_0^1 P dR \tag{16}$$

3.2 DL Training Results

The initial results for the training of the DL networks are shown in Table 3. Both networks achieve a high level of performance, with Visible network achieving the highest performance in precision while thermal only network achieves the highest performance in recall and average precision.

Table 3 Results of initial training of DL networks on synthetic dataset, the best score in each metric is indicated with Bold text.

DL Network	Precision	Recall	Average Precision
Visible	**0.942**	0.922	0.848
Thermal	0.932	**0.966**	**0.872**

3.3 Visual Odometry Trajectories

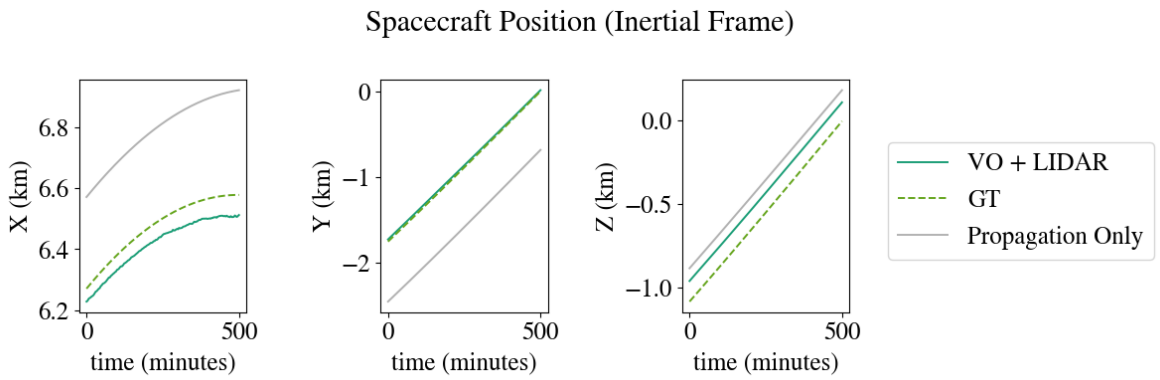

Fig. 8 The spacecraft's relative position to the asteroid in the inertial frame, r_{sc}, shown by the GT trajectory and the trajectories estimated from Visible image VO with LIDAR range measurements and estimated from propagation only.

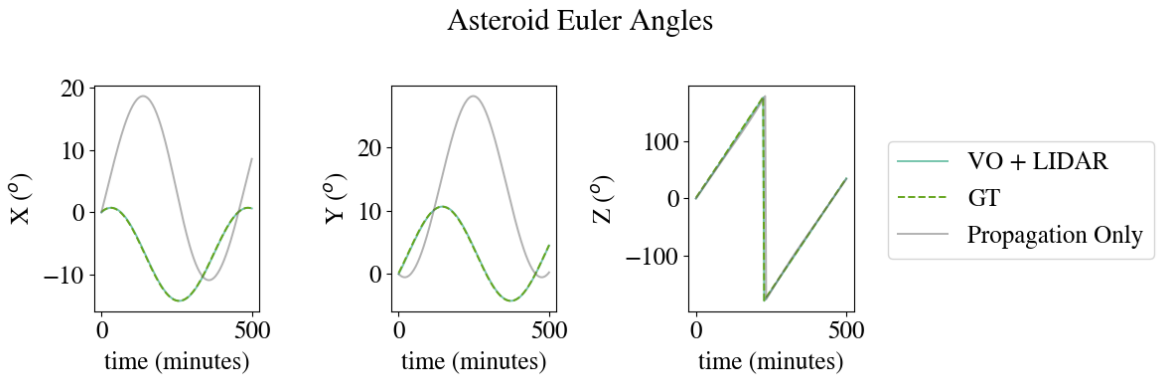

Fig. 9 The Euler Angles for the Asteroid's attitude, R_A, shown by the GT and the Euler angles estimated from Visible image VO with LIDAR range measurements and estimated from propagation only.

To assess the performance of the VO using the factor graph and the DL keypoint matches, two trajectories were generated with a semi-major axis of 6.5km, the first trajectory with a low sun phase angle, and the second trajectory with a high sun phase angle. PANGU was used to generate the visible and thermal images using the parameters of the visible camera, ONC-T, and thermal camera, TIRI, both of which are instruments on Hayabusa2. The spacecraft's estimated position for the first generated trajectory is shown in Figure 8 and the asteroid's attitude is shown in Figure 9. The VO is achieved using the visible keypoint matching and the LIDAR range measurement, and is compared to the Ground Truth (GT), and the propagated trajectory from the initial state provided to the factor graph. The factor graph is initialised with a 10% error in the spacecraft's position, 0.1% in the magnitude of the angular velocity, and a 15^o error in the angle of rotation. It can be clearly seen that the Visible image based VO method significantly improves the position and attitude estimation over propagation from the initial estimate only.

3.4 Visual Odometry Results

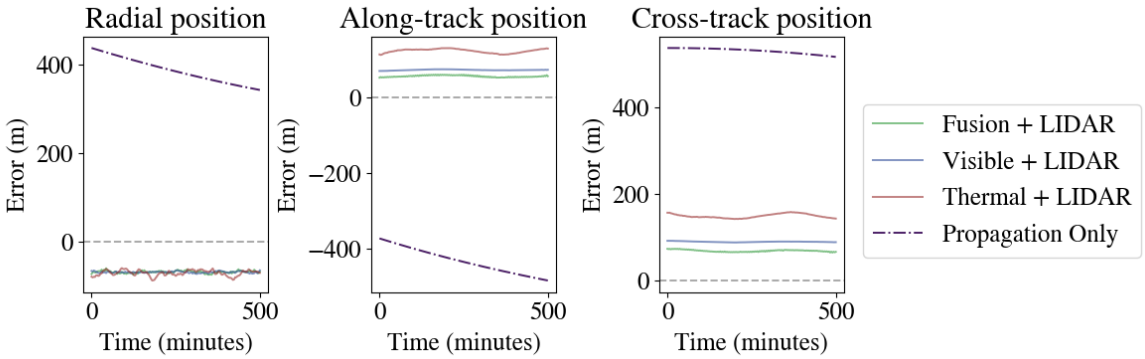

Fig. 10 Error in position estimation in LHLV frame showing the performance VO with LIDAR range measurement using Visible, Thermal, and Fusion keypoint matching relative to propagation only. The trajectory used has a low sun phase angle.

The performance of the three different keypoint matching methods — visible only, thermal only, and fusion— used for VO are shown in Figures 10 and 11. Lidar range measurements are also used in the factor graph to remove ambiguity in scale. In figure 10 the errors in the positions relative to the ground truth are shown in the Local Horizontal Local Vertical Frame (LHLV). The methods can all be seen to achieve a similar level of accuracy in the radial direction. In the along track and cross track directions fusion performs the best, with visible only matching achieving slightly lower performance, and thermal only performing worst, with an error almost double the size of fusion.

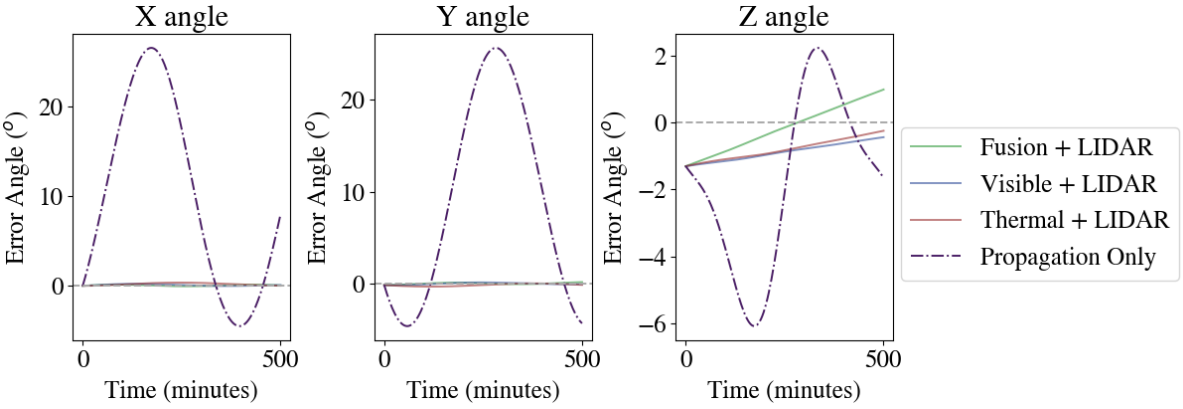

Fig. 11 Error in the Euler Angles of the estimated asteroid's attitude relative to the inertial reference frame showing the performance of VO with LIDAR range measurement using Visible, Thermal, and Fusion keypoint matching relative to propagation only. The trajectory used has a low sun phase angle.

In figure 11 the error in the Euler angles is shown between the 3 different matching types and the propagation only method. The VO method significantly improves the estimation of the asteroid's attitude in the X and Y Euler angles (the directions with the most significant initial error) achieving $< 1^o$ error from an initial estimate that has $> 20^o$ error. In the Z direction, the VO methods also improves the attitude estimation, although by a relatively smaller amount.

The results for the second high sun phase angle trajectory are shown in Figures 12 and 13. Figure 12 shows to position error in LHLV where it can be seen all the methods perform significantly worse than for the low sun phase angle trajectory, with errors in the range of 200-400m in the cross track and along track directions. The error also increases in the estimation of the asteroid's attitude, most significantly when using visible only keypoint matching, while thermal only and fusion based matching achieve similar levels of performance, although still significantly worse than the low sun phase angle performance.

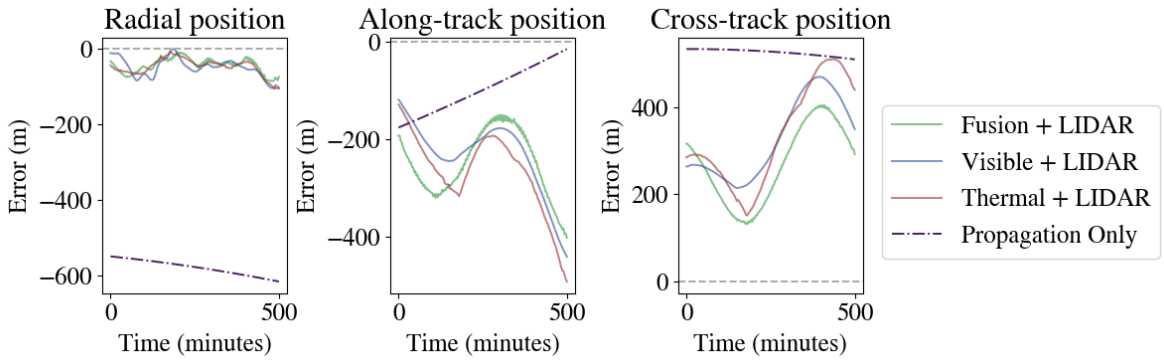

Fig. 12 Error in position estimation in LHLV frame showing the performance VO with LIDAR range measurement using Visible, Thermal, and Fusion keypoint matching relative to propagation only. The trajectory used has a high sun phase angle.

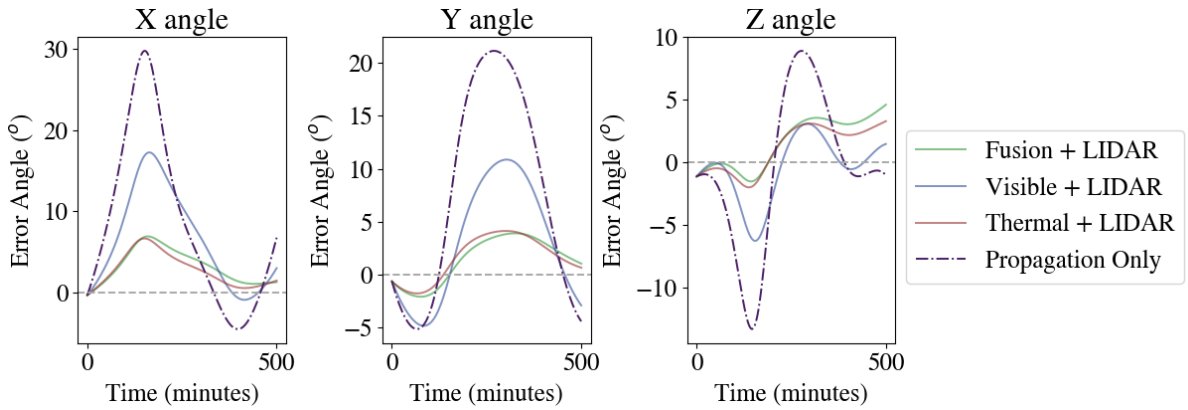

Fig. 13 Error in the Euler Angles of the estimated asteroid's attitude relative to the inertial reference frame showing the performance of VO with LIDAR range measurement using Visible, Thermal, and Fusion keypoint matching relative to propagation only. The trajectory used has a high sun phase angle.

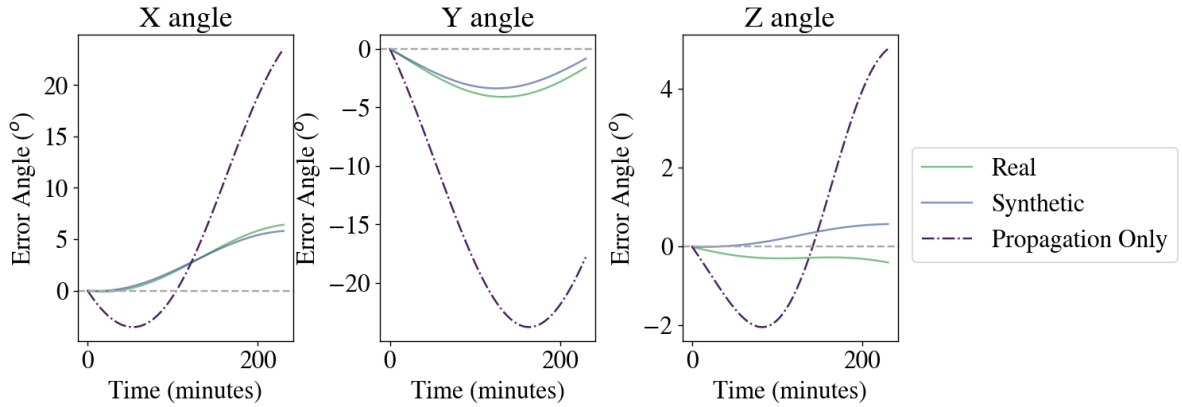

Fig. 14 Error in the Euler Angles of the Ryugu's estimated attitude relative to the inertial reference frame showing the performance of Visible image VO with LIDAR range measurement on Real and Synthetic images of Ryugu.

Real images from the spacecraft Hayabusa2 are used to validate the visivle DL model and VO method. Real images are gathered from NASA's Planetary data system [19], while matching synthetic images are generated from the trajectory using PANGU. The trajectory was found using the Hayabusa2 SPICE kernels[23] and a 0.1% in the magnitude of the angular velocity, and a 15^o error in the angle of rotation is introduced to reflect the initial uncertainties in the rotational state of Ryugu[24]. Figure 14 shows the errors in the Euler angles of the attitudes estimates using visible and synthetic images.

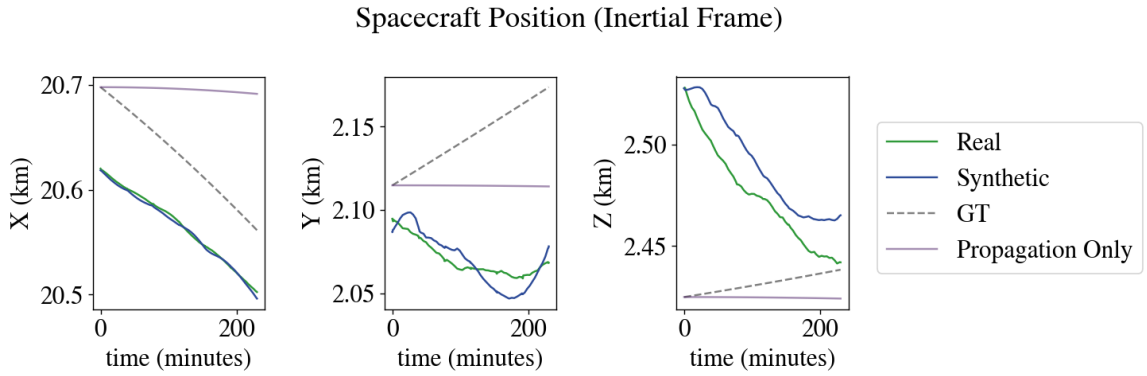

Fig. 15 GT and estimated spacecraft position using visible only VO on real and synthetic images.

The method used for navigation is VO with LIDAR and visible image only keypoint matching. The VO method significantly improves the attitude estimation relative to the propagation only method, achieving a $< 6^o$ error in all Euler angles. The VO method achieves similar levels of performance on the real and synthetic images, demonstrating that the DL network is able to adapt between real and synthetic images. Figure 15 shows the spacecraft's GT and estimated position in the inertial reference frame, where no initial position error is introduced. The position error of the VO method increases for Real and Synthetic images, and for propagation only showing that there are limitations in the performance of the method as it is applied to real trajectories.

4 Discussion

The VO method is shown to be able to effectively track surface features and estimate the pose of the spacecraft for a low sun phase angle trajectory. From an initial estimate with a large error the method is able to correct the position estimation error to within $< 5\%$ of the range, and the error in the asteroid's attitude to $< 2^o$. This demonstrates that DL is able to provide sufficiently accurate keypoint matches for pose estimation, and that a factor graph based VO formulation is able to effectively adapt to a large initial error. When tested on a higher sun phase angle trajectory, in which illumination conditions are worse, the VO has a larger error, especially for visible only keypoint matching. Thermal and fusion based VO are more robust to adverse illumination conditions, but there is still a drop in performance when using thermal and fusion VO at high sun phase angles showing that they are still affected by illumination conditions, so additional work must be completed to improve the performance or operations limited to only high sun phase angle trajectories. Fusion based matching doesn't appear to significantly improve the performance of pose estimation over visible VO at low sun phase angles or improve over thermal VO at high sun phase angles, but it does achieve relatively low errors in both cases, where visible and thermal only perform best in one case. This shows that fusion VO achieves the aim of providing the best of both the individual sensor types.

The use of the VO method for estimating Ryugu's attitude using synthetic and real images results show that the visible DL network trained on PANGU images is able to adapt to real images, demonstrated by the small difference in performance on real and synthetic images. This is primarily due to the PANGU images generated for training being highly representative of the real asteroid images. The performance of VO still drops when used on the real trajectory, but because it is demonstrated this is not due to the DL networks, it instead must arise from differences in the trajectory, unmodeled dynamics, and image capture rates.

Errors in the radial position estimation remain similar and fairly small, primarily determined by the LIDAR range measurement, while the errors in position estimation in the along track and cross track

direction can get significantly larger, likely arising from the landmarks not being directly constrained to the correct distance from the asteroid's centre, but instead their position is inferred from the spacecraft's and asteroid's motion. To address the cross-track and along-track errors the use of a centroid measurement could be considered, which directly constrain spacecraft position, or the use of loop closure methods to constrain the positions when the spacecraft revisits previous locations over the asteroid.

5 Conclusions

In this work we have demonstrated the use of DL for matching keypoints between images of an asteroid combined with a factor graph based visual odometry method estimate the relative pose of a spacecraft to an asteroid. We are able to use the factor graph to effectively fuse visible and thermal keypoint measurements, as well as fusing LIDAR range measurements. We show that the use of thermal images and sensor fusion improve the performance of navigation under adverse illumination conditions. We have shown that the VO method can improve the attitude estimation on real asteroid images, demonstrating the ability of the visible DL model to generalise to real images.

Future work on the development of the factor graph should explore the use of loop closure and the use of the Incremental Smoothing And Mapping (ISAM) engine available through GTSAM. Loop closure will improve the pose estimation by reducing drift over long periods achieving a full SLAM method, while ISAM will allow for faster execution of the method allowing for online implementation. The use of real thermal images from Hayabusa2 should be used to test the Thermal and Fusion VO methods on real trajectories, to demonstrate the generalisability of the thermal DL model. Additionally a larger number of trajectories, both real and synthetic, should be used to test the VO method to help characterise its performance under different conditions.

Acknowledgments

We would like to acknowledge the Strathclyde doctoral training centre for AI based multi-sensor data fusion for space applications for co-funding this work funding this work.

This work is supported by the following ESA Open Space Innovation Platform grant:AI-based multi-sensor data fusion for autonomous navigation around asteroids, ESA contract number: I-2023-05255

Declaration of Use of Artificial Intelligence

AI is used in the form of the Deep Learning networks used in keypoint extraction and matching in the method. Additionally to search for specific functions in Python libraries Claude 4.6 sonnet was used. Once the functions were found they were always checked against the libraries documentation to ensure they behaved as expected. Outside of these two purposes no AI was used in this work.

References

[1] Go ONO, Fuyuto TERUI, Naoko Ogawa, Yuya MIMASU, Kent Yoshikawa, Seiji YASUDA, Kota MATSUSHIMA, Yuto Takei, Takanao SAIKI, and Yuichi Tsuda. Gnc design and evaluation of hayabusa2 descent operations. *TRANSACTIONS OF THE JAPAN SOCIETY FOR AERONAUTICAL AND SPACE SCIENCES, AEROSPACE TECHNOLOGY JAPAN*, 19:259–265, 03 2021. doi: 10.2322/tastj.19.259.

[2] David A. Lorenz, Ryan Olds, Alexander May, Courtney Mario, Mark E. Perry, Eric E. Palmer, and Michael Daly. Lessons learned from osiris-rex autonomous navigation using natural feature tracking. In *2017 IEEE Aerospace Conference*, pages 1–12, 2017. doi: 10.1109/AERO.2017.7943684.

[3] Carmine Buonagura, Mattia Pugliatti, and Francesco Topputo. Image processing robustness assessment of small-body shapes. *The Journal of the Astronautical Sciences*, 69, 11 2022. doi: 10.1007/s40295-022-00348-6.

[4] Carlo Tomasi. Detection and tracking of point features. volume 9, page 3, Norwell, MA :, 1991. Kluwer Academic Publishers. `https://api.semanticscholar.org/CorpusID:238434334`.

[5] A Pellacani, M Graziano, M Fittock, J Gil, and I Carnelli. Hera vision based gnc and autonomy. In *Proceedings of the 8th European Conference for Aeronautics and Space Sciences. Madrid, Spain, 1-4 july 2019*, 2019. doi: 10.13009/EUCASS2019-39.

[6] Leiyu Chen, Shaobo Li, Qiang Bai, Jing Yang, Sanlong Jiang, and Yanming Miao. Review of image classification algorithms based on convolutional neural networks. *Remote Sensing*, 13(22), 2021. Cited by: 282; All Open Access, Gold Open Access. doi: 10.3390/rs13224712.

[7] Yun Liao, Yide Di, Kaijun Zhu, Hao Zhou, Mingyu Lu, Yijia Zhang, Qing Duan, and Junhui Liu. Local feature matching from detector-based to detector-free: a survey. *Applied Intelligence*, 54(5):3954–3989, 2024. doi: 10.1007/s10489-024-05330-3.

[8] Travis Driver, Katherine A. Skinner, Mehregan Dor, and Panagiotis Tsiotras. Astrovision: Towards autonomous feature detection and description for missions to small bodies using deep learning. *Acta Astronautica*, 210:393–410, 2023. ISSN: 0094-5765. doi: https://doi.org/10.1016/j.actaastro.2023.01.009.

[9] Olli Knuuttila, Antti Kestilä, and Esa Kallio. Cnn-based local features for navigation near an asteroid. *IEEE Access*, 12:16652–16672, 2024. doi: 10.1109/ACCESS.2024.3358021.

[10] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: an efficient alternative to sift or surf. pages 2564–2571, 11 2011. doi: 10.1109/ICCV.2011.6126544.

[11] D.G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157 vol.2, 1999. doi: 10.1109/ICCV.1999.790410.

[12] Andréa Macario Barros, Maugan Michel, Yoann Moline, Gwenolé Corre, and Frédérick Carrel. A comprehensive survey of visual slam algorithms. *Robotics*, 11(1), 2022. ISSN: 2218-6581. doi: 10.3390/robotics11010024.

[13] B. Razgus, E. Mooij, and D. Choukroun. Relative navigation in asteroid missions using dual quaternion filtering. *Journal of Guidance, Control, and Dynamics*, 40(9):2151–2166, 2017. doi: 10.2514/1.G002805.

[14] Xi Ma, S p Gong, and Lin Cheng. Autonomous vision-based control for spacecraft navigation around an asteroid. *IEEE Transactions on Aerospace and Electronic Systems*, PP:1–14, 01 2025. doi: 10.1109/TAES.2025.3561730.

[15] Mehregan Dor, Travis Driver, Kenneth Getzandanner, and Panagiotis Tsiotras. Astroslam: Autonomous monocular navigation in the vicinity of a celestial small body—theory and experiments. *The International Journal of Robotics Research*, 43(11):1770–1808, 2024. doi: 10.1177/02783649241234367.

[16] Federico Castanedo. A review of data fusion techniques. *The Scientific World Journal*, 2013, 2013. ISSN: 1537744X. doi: 10.1155/2013/704504.

[17] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 337–33712, 2018. doi: 10.1109/CVPRW.2018.00060.

[18] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. Lightglue: Local feature matching at light speed. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 17581–17592, 2023. doi: 10.1109/ICCV51070.2023.01616.

[19] Murakami Shin ya; Yamamoto Yukio; Crombie M. Katherine; Ishihara Yoshiaki. Hayabusa2 optical navigation camera (onc) bundle. PDS Asteroid/Dust Subnode, 2023. doi: 10.17597/isas.darts/hyb2-00200, https://sbnarchive.psi.edu/pds4/hayabusa2/hyb2_onc/.

[20] F. Dellaert and M. Kaess. Factor graphs for robot perception. *Foundations and Trends in Robotics, FNT*, 6(1-2):1–139, Aug. 2017. doi: 10.1561/2300000043.

[21] Frank Dellaert and GTSAM Contributors. borglab/gtsam, May 2022. doi: 10.5281/zenodo.5794541, https://github.com/borglab/gtsam).

[22] Peter J Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution*, pages 492–518. Springer, 1992.

[23] S Murakami, Y Yamamoto, and Y Ishihara. Hayabusa2 spice kernel archive bundle. *JAXA Data Archives and Transmission System*, page 3, 2022. doi: 10.17597/isas.darts/hyb2-00600.

[24] Koji Wada, Matthias Grott, Patrick Michel, Kevin J. Walsh, Antonella M. Barucci, Jens Biele, Jürgen Blum, Carolyn M. Ernst, Jan Thimo Grundmann, Bastian Gundlach, Axel Hagermann, Maximilian Hamm, Martin Jutzi, Myung-Jin Kim, Ekkehard Kührt, Lucille Le Corre, Guy Libourel, Roy Lichtenheldt, Alessandro Maturilli, Scott R. Messenger, Tatsuhiro Michikami, Hideaki Miyamoto, Stefano Mottola, Thomas Müller, Akiko M. Nakamura, Larry R. Nittler, Kazunori Ogawa, Tatsuaki Okada, Ernesto Palomba, Naoya Sakatani, Stefan E. Schröder, Hiroki Senshu, Driss Takir, Michael E. Zolensky, and International Regolith Science Group (IRSG) in Hayabusa2 project. Asteroid Ryugu before the Hayabusa2 encounter. *Progress in Earth and Planetary Science*, 5(1):82, Dec. 2018. ISSN: 2197-4284. doi: 10.1186/s40645-018-0237-y.