Vision Technologies for Small Body Proximity Operations

Adnan Ansar*, Yang Cheng*

*Computer Vision Group, Jet Propulsion Laboratory,
California Institute of Technology, USA
e-mail: {ansar,ycheng}@jpl.nasa.gov

Abstract

We present a framework for use of computer vision technologies to localize a spacecraft during small body (comets and asteroids) proximity operations. Our approach is to first detect image-based landmarks at stand-off distance during a mission survey phase, then catalog these landmarks into an easily referenced database, and finally use the cataloged data to recognize the landmarks during proximity operations. The catalog includes 3D body relative coordinates for the landmarks, so that image derived bearing angles to the landmarks enable localization of the spacecraft.

In this paper, we describe our method of landmark detection and recognition, the details of the landmark catalog including estimation of 3D body-relative landmark locations, and our approach to vision-based estimation of spacecraft pose (position and attitude). We validate our research using real data from the JAXA MUSES-C mission and the NASA Deep Impact and NEAR missions as well as through detailed simulations.

1 Introduction

Any mission, such as sample return, requiring operations in close proximity to a small body demands a high degree of autonomy. This imposes a requirement for precise *in-situ* spacecraft localization with respect to the target body. A vision sensor provides a low cost, low weight, low power, flight proven solution. We present a suite of computer vision technologies to enable vision-based spacecraft localization.

Our operational scenario is a mission with one or more survey phases at a stand-off distance from the target body. During this stage, we detect a set of landmarks specifically chosen to be recognizable at subsequent stages of the mission. For this research we use a variant of the Scale Invariant Feature Transform (SIFT) [7]. SIFT detects highly salient points in the image and associates to each a descriptor based on local gradient data. These points are our landmarks. The descriptor is used to recognize previously detected landmarks when encountered again. The descriptors are highly invariant to image scale change and in-plane rotation, thus accommodating changes in spacecraft location and attitude relative to the target body. Illumination invariance is achieved through a combination of image processing techniques and incorporation of sun angle. Details of the landmark detection and recognition are given in Section 2.1.

During the survey phase, detected landmarks are tracked across multiple image frames using image correlation techniques. These tracks are use to generate 3D locations for the landmarks in a body-centric coordinate frame using a least-squares optimization approach commonly referred to as Bundle Adjustment. The landmarks are cataloged by descriptor, 3D location and other contextual information. The catalog is designed for robustness, a high degree of discriminability between landmarks, and ease of search. Details of the catalog generation process are given in Section 2.2. While the BA process produces very accurate spacecraft pose and landmark positions, it is a computationally expensive, strictly batch process. Thus, it is ideal for generating the landmark catalog from stand-off distances but not for proximity operations, which require fast computation.

During proximity operations, landmarks are detected and their descriptors compared to those recorded in the catalog. The combination of image coordinates and previously extracted 3D catalog locations allows for a complete body-relative 6 DoF solution for the pose of the spacecraft [3][8]. Note that in an actual mission, the 2D-3D correspondence between image coordinates and 3D locations would be incorporated directly into the navigation filter rather than through an intermediate, purely vision-based pose solution. However, the vision-based pose is still critical as both a sanity check on the filter and as an outlier rejection mechanism for catalog match errors prior to handing off data to the filter. In this paper, we focus on the vision derived pose only. A description of how the vision products can be incorporated into a navigation filter is given in [4].

We begin with an overview of the vision technologies used. We then show results on both real and synthetic data to validate the utility of this approach.

2 Algorithm Overview

We now describe the vision algorithms developed and used for this research.

2.1 Landmark Detection and Recognition

While our overall approach is agnostic to the choice of landmark type, we chose SIFT as a starting point because

of prior success in the vision literature and its inherent scale and rotation invariance. A detailed overview of the base algorithm is beyond the scope of this paper but can be found in [7]. In brief, SIFT selects keypoints based on response to an approximation of the Laplacian of Gaussian (LoG) filter applied to a scale-space representation of an image. Here, scale-space refers to a pyramid-like hierarchy consisting of multiple copies of the image, successively downsampled and convolved with a series of spatial band-pass filters. Points with local maximum response across both the spatial and scale directions are chosen as landmarks. Thus, the resulting landmarks are not only optimally localized in the image but chosen at the most optimal scale. When viewed from a different distance, a given landmark will have a different size in the image but will be processed at a scale consistent with the prior view. Note that while landmarks are distinctive from an image processing standpoint, they may not correspond to obviously recognizable topographic features.

Landmarks are identified via a descriptor consisting of local gradient data. For each landmark, a series of oriented gradient histograms is computed in windows surround the landmark center. These are concatenated into a single 128 element vector that serves as the landmarks descriptor. Matching of two landmarks is accomplished by computing the Euclidean distance between their descriptors. The histograms are computed only at the optimal image scale. Hence, the descriptor formulation is inherently scale invariant. Prior to descriptor computation, the region around each landmark is rotated so that the dominant gradient direction is up, thus providing rotation invariance in the image plane. Further, the use of multiple windows for computation of gradient histograms instead of a single large window affords some degree of shift invariance.

Figure 1 shows a landmark detection and matching result for two images of asteroid Eros acquired during the NASA Near Earth Asteroid Rendezvous (NEAR) mission.

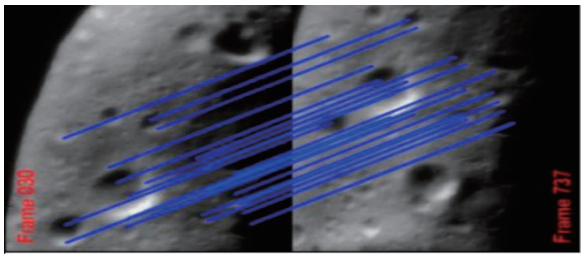

Figure 1: Detection and recognition of landmarks from images of two different orbits of Eros taken by the NEAR spacecraft.

The images are from different orbits and no frame-to-frame tracking is involved in obtaining landmark correspondences. Rather, this is a process in which previously observed landmarks are reacquired and recognized via their descriptors. Hence, we are treating landmark matching as a recognition problem rather than as a tracking problem.

The work presented in this paper used our own variant implementation of SIFT, which diverges from [7] in some of the finer points including the details of the scale-space construction and some image pre-processing.

While SIFT has been highly successful in general settings, we found that some modifications were required for the high contrast and sharp illumination transitions typical of space based imagery of small bodies. When used directly, there were too many local maxima in the LoG response, resulting in unstable landmarks. We experimented with various spatial image filters to improve response. If $I(x,y)$ is the intensity of the image at pixel location (x,y), then we are interested in filters of the type

$$\hat{I}(x,y) = I(x,y) - B * I(x,y) \qquad (1)$$

where B is an operator approximating a spatial low-pass filter. Thus, the resulting $\hat{I}$ is an approximation of a high-pass filtered version of I.

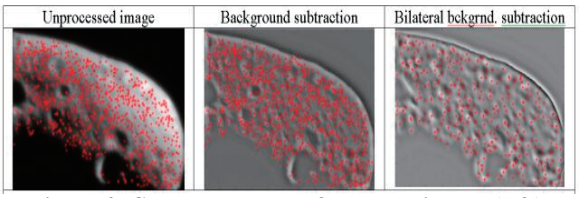

Figure 2: SIFT landmarks from raw image (left), high-pass (center), and bilateral high-pass (left).

Figure 2 shows SIFT features on a raw image of Eros (left) as well as using two different choices of B. The first is convolution with an averaging kernel (center) and the second is the bilateral filter (right). While both high-pass filter operations produced more stable landmarks with a higher likelihood of recognition in later imagery, the bilateral high-pass [2] had the most dramatic improvement over the widest collection of datasets, both real and synthetic. Subjectively, one observes that the number of landmarks returned after bilateral processing on the right side of Figure 2 is smaller than for the other two cases. These fewer landmarks, however, are better candidates for recognition in later imagery.

Scale invariance in SIFT is a function of image resolution and the errors introduced in downsampling during scale-space construction, but it is generally restricted to a factor of 2 or 3. For a mission using both a wide and narrow angle camera and multiple staging/survey phases, we find that it is possible to extend this scale invariance to a factor of well over 10 in a boot-strap fashion by associating descriptors across

discrete steps. This data association can be incorporated directly into the landmark catalog by assigning the same landmark identification to descriptors derived at various distances. In Figure 3 we show an example of this sort of descriptor chaining on imagery of asteroid Tempel 1 acquired by the NASA Deep Impact mission. A specific landmark (center pixel of image 3.D) is chosen in the closest image and detected in the next closes (3.C) by descriptor matching. This is indicated by the overlay in 3.G. This is chained together in pairwise fashion with increasingly distant imagery until the landmark can be localized at the most distant scale (3.H). The end result is an ability to match the given landmark across an image scale change of 17x, giving global context even at very close proximity to the surface.

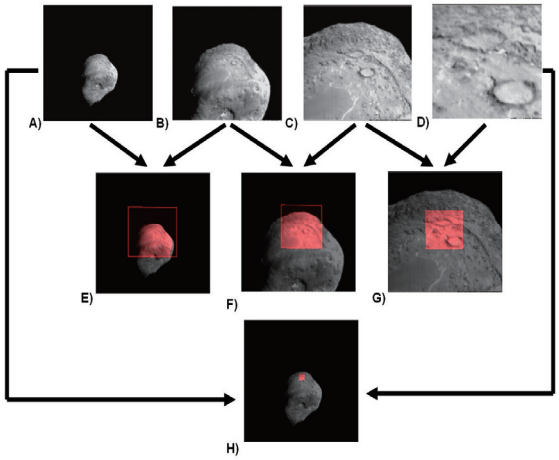

Figure 3: Chaining descriptors across scales to achieve invariance across wide change in distance. Above shows matching across 17x scale change.

The standard SIFT algorithm is especially well suited to detecting and recognizing features on flat surfaces. In the presence of high local surface relief, a likely scenario for a spacecraft near a small body, the matcher quickly breaks down. Our workaround is to extract landmarks only in areas that are relatively locally flat. We can compute local surface roughness either from binocular stereo on adjacent frames or from multi-baseline stereo using pose from the BA algorithm described in section 2.2.

Figure 4 shows a synthetic image (left), a stereo reconstruction (center) and the roughness map (right) derived from stereo. We threshold the roughness estimate to select landmarks only in flat areas likely to be recognizable from multiple viewpoints. In practice, we use a more sophisticated, highly efficient approach that reduces the problem of surface topography estimation to a 1-D line search in the neighborhood of a landmark [5].

An additional enhancement to the basic SIFT framework is to incorporate information on body-relative location, sun angle and other metadata directly into the catalog. This restricts the search space for descriptor matching and eliminates many outliers.

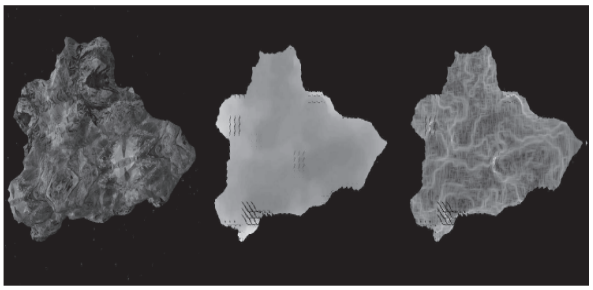

Figure 4: Synthetic image (left), stereo map (center), and roughness map (right). Roughness used to prune landmarks in unstable areas.

2.2 Generating 3D Landmark Positions

Detected landmarks are tracked across adjacent frames. Since the imagery in question is closely grouped in time, viewpoint and illumination are similar. Hence, standard image correlation techniques are employed [10]. At least 2 frames of tracking data are required to recover 3D. We also extract other cues such as surface normal and local roughness to better prune the most stable landmarks for subsequent detection.

3D landmark locations are derived from bundle adjustment (BA), a technique for simultaneously refining the position of landmarks as well as the pose of all camera frames involved. The approach was originally conceived in the field of photogrammetry and has been used to analyze airborne photographic imagery for many years. Recently, BA has been increasingly used by vision researchers and has been widely recognized as one of the most reliable and accurate techniques for 3D scene reconstruction. Assume that n 3D points ($a_j, j = 1, 2, \ldots n$) are observed in m views (b_i, $i = 1, 2, \ldots m$) and let X_{ij} be the observed projection of the i-th point on image j. Given the initial locations of these 3D points and the initial camera extrinsic parameters (i.e. pose data), these 3D points can be reprojected to the m image planes following the known imaging geometry of the camera. Ideally, the reprojected image points should match exactly the landmark points observed in the images. In reality, due to numerous uncertainties such as initialization error and image noise, the reprojected points do not match the observed points. The main objective of BA is to minimize the reprojection error with respect to all 3D points and the camera's extrinsic parameters, specifically

$$\arg\min_{aj,bj} \sum_{i=1}^{n} \sum_{j=1}^{m} d(Q(a_j, b_i), X_{ij})^2 \qquad (2)$$

where $Q(a_j, b_i)$ is the predicted projection of point i on image j and $d(x, y)$ denotes the Euclidean distance between image point x and y.

The BA method can be cast as a non-linear optimization problem. If the dimension of each a_j and b_i is κ and λ respectively, the total number of parameter in a BA system is $m\kappa+n\lambda$. This can result in a very large system. Fortunately, the lack of interaction among parameters for different 3D points and cameras in multiple views results in a very sparse Jacobian matrix with independent blocks. A sparse matrix manipulation technique, typically an iterative method, can be used to gain considerable computational speedup. We used a generic bundle adjustment based on a sparse implementation of the Levenberg-Marquardt algorithm by Lourakis and Argyros [11].

We use the first camera frame as the global coordinate system, and the rest of the imagery is referenced to this first frame. Once the 3D point cloud associated with the landmarks is computed, we transfer the results to a body-centric coordinate frame by centroiding on the landmarks. For the very first pair of images, we use a linear closed form solution of Structure from Motion (SFM) to obtain the relative motion [9]. SMF is used only in the first pair of images. The rest of the image poses are determined initially by a pose estimation algorithm using [1][8]. BA proceeds iteratively across groups of frames. In summary, poses for new frames are established from known 3D landmark locations computed in previous BA steps. Then the 3D positions of fresh landmarks are determined by triangulation. Landmark position uncertainty is given by covariance analysis derived from the BA procedure [5].

Note that BA does not recover a global scale, since scale is inherently ambiguous in image data. However, in practice we can impose a scale constraint by using the navigation filter to determine the distance between the first two frames used for BA.

Once the spacecraft pose and landmark positions are determined, we use multi-baseline stereo methods [12] to determine the local topography surrounding each landmark. Then the landmark surface normal and roughness are estimated by using the extracted topography.

2.3 Catalog Components

In order to enable robust and efficient matching, the catalog entry for each landmark must be a comprehensive yet compact snapshot of its properties. We show in Table 1 the various components that make up a complete landmark database.

The two key components here are the landmark position and descriptor. Under ideal conditions, these suffice to completely solve the spacecraft localization problem, since the descriptor is used for landmark recognition in subsequent images, and the pose is given by the 3D landmark positions and their corresponding 2D image coordinates. However, several other considerations go into the catalog. For both efficient query and to minimize false matches, we consider only landmarks detected under similar viewpoint and sun angles (sensor and sun vector). Covariance estimates as well as normal and roughness data provide a confidence metric on a given landmark and can be used both in matching and directly by a state estimator in the case of the covariance. Finally, maintaining context (topography and tracks) is useful for both outlier rejection in the match process and whenever augmentation of the landmark catalog with new entries is desired, such as in a mission with a multi-stage survey phase.

Table 1: Components of Landmark (LM) Database

LM parameters	Description
Position	3D position in body frame
Descriptor	SIFT (or other) descriptor
Position covariance	6 elements of symmetric 3x3 matrix
Normal	Unit vector in body frame indicating surface normal
Roughness	Variation in surface elevation
Sensor vector	Vector from body frame to sensor
Sun vector	Unit vector to sun in body frame
Topography	8 3D points surrounding the landmark
Landmark track	Landmark image location in next three frames
Landmark size	Image scale of landmark (i.e. location in scale-space representation)

2.4 Localization from Landmarks

After the landmark catalog is generated during the mission survey phase, it is used during proximity operations to localize the spacecraft relative to the target body. Assume initially that descriptor matching suffices to provide an error-free correspondence between observed landmarks and their catalog entries. This means that we have a direct correspondence between 3D points, expressed in the coordinate frame of the small body, and their 2D image coordinates. This correspondence is sufficient in the general case to solve for the 6 DoF position and attitude of the camera in the small body frame. If the camera is fully calibrated, there exists a known function F, such that for any camera pose (R,T) in the small body coordinate frame and a 3D point w expressed in the same frame, the corresponding image coordinate p is given by

$$p = F(R,T,w) \tag{3}$$

Let $\{w_i\}$ be a collection of 3D points corresponding to landmark catalog entries, and $\{p_i\}$ the corresponding image coordinates from a frame acquired in proximity of the small body. For a hypothetical camera pose (R',T') let

$$e(p_i, w_i, R', T') = N(p_i - F(R', T', w_i)) \quad (4)$$

be the associated image reprojection error, where N is an appropriately chosen norm. Solving for pose then amounts to finding

$$(R,T) = \arg\min_{R',T'} \sum_i e(p_i, w_i, R', T') \quad (5)$$

Details on the efficient solution to Eqn. (5) as well as studies on robustness and potential singularities can be found in [3][8].

As mentioned previously, in a real mission scenario, the $\{w_i, p_i\}$ correspondence data would be fed directly to the navigation filter. However, the pose solution from Eqn. (5) is critical as both a sanity check on the system and as an outlier rejection scheme for poor landmark matches or complete mismatches. In particular, the navigation filter can accommodate poor localization (larger covariance) but not outright match failures. Hence, we must eliminate these failures using only vision techniques.

Observe that the reprojection error in Eqn. (4) gives a measure of consistency of a given point with the overall pose solution. A naïve outlier rejection scheme would be to threshold directly on the reprojection error. However, this ignores the fact that the pose solution is partially conditioned on any mismatches in the data. Our solution is to use the Random Sample Consensus (RANSAC) [6] framework for outlier rejection. Assume K total matches in the database. We outline the procedure below:

- For m iterations:
  - Randomly select n landmarks matches from the K, where $n << K$, and solve for (R,T) in Eqn. (5) using only these matches
  - Compute the median error from Eqn. (4) over all K points
  - If median error is smaller than previous smallest median error, save current (R, T) as best tentative model
- Compute error for all K matches using best (R, T)
- Threshold to reject outliers
- Recompute (R, T) using only inliers

In the above, m is derived from a principled statistical argument [6] based on estimates of outlier probability and on requirements for algorithm correctness. Figure 5 shows the outlier rejection mechanism in action on a synthetic scene. In this case, the three outliers (red) were intentionally introduced and automatically detected.

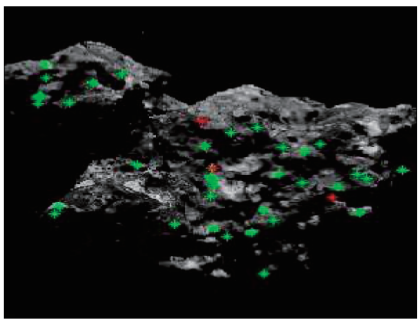

Figure 5: Detected and recognized landmarks in green. Intentionally introduced outliers in red are found and rejected.

3 Experimental Validation

We have already shown some experimental results on matching (Figure 1 and Figure 5) and scale invariance (Figure 3) in the course of the algorithm description. In this section, we focus on two more detailed case studies. The first addresses localization accuracy using a synthetic object. The second focuses on both catalog generation and accuracy using a model of Itokawa as the target body.

3.1 Localization: Twice around case study

We assemble the pieces of this research effort into an experiment called the "twice around" study. This combines landmark detection, a simplified version of catalog generation, and spacecraft localization from matched landmarks. The name refers to flying twice around a small body. In the first orbit, landmarks are detected and stored in a catalog along with descriptors and 3D locations. In the second orbit, landmarks are detected and matched via their descriptors to catalog entries. The recorded 3D landmark locations from the first orbit are used with image coordinates from the second orbit to localize the spacecraft using image based pose estimation.

For this study, we adopt a simplified version of the landmark catalog that uses only 3D locations and descriptor data. Rather than using the full BA framework, this limited catalog uses trinocular stereo from triplets of neighboring frames to localize the landmarks.

The target body is a purely synthetic object (See Figure 5) approximately 500m in diameter, and the spacecraft orbits at approximately 2km from its center of mass. Two hundred frames are used in the first orbit to generate a landmark catalog. During the second orbit, 70 evenly spaced frames are used to evaluate localization against ground truth. Note that the two orbits are not identical and that the imaging positions change between the two passes.

Landmark detection during the second orbit uses the

RANSAC approach described above for robustness. In Figure 6, we show one frame with landmarks that were matched to the catalog as well as the vision based pose estimate for the 70 frames of the second orbit. Black dots indicate the vision estimate of spacecraft position, while the green coordinate frames indicate ground truth. The cyan points show all entries in the landmark catalog.

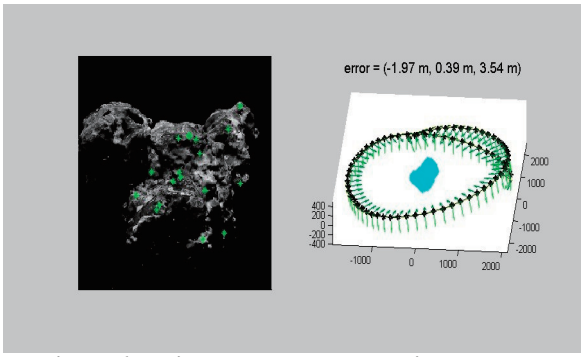

Figure 6: Twice around study showing matched landmarks (left) in one frame and vision-based pose estimate (right) over 70 frames of second orbit.

Error covariance is computed in the camera frame rather than the body frame for consistency across images. We find that the 1σ error in the principal direction (nearly aligned with the optical axis of the camera) is 5.7m or approximately 0.4% of the distance to the surface of the body. The absolute position error is bounded by 23m over the course of the second orbit and has an RMS value of 8m over all frames. Note that while good, these results are purely vision-based. Incorporation into a navigation filter is expected to improve them dramatically.

3.2 Bundle Adjustment and Landmark Catalog: Twice down case study

For this study, we use a shape model of asteroid Itokawa developed through the JAXA MUSES-C mission. Texture is added to the shape model, and virtual imagery is rendered from this data based on simulated spacecraft position and attitude.

The virtual spacecraft captures imagery while performing the following maneuvers twice (hence the name of the study) with some random variation in trajectory:

- Start at 300m above the target
- Hover within circle of radius 30m
- Descend to 150m
- Hover within circle of radius 30m

The first descent is used to generate a full catalog following sections 2.2 and 2.3. The second descent is used to validate localization from the catalog.

After the first descent, the BA algorithm estimates spacecraft pose and 3D landmark locations. For each landmark, the input includes its pixel location in a base image as well as three successive images, and the size of the image patch corresponding to the landmark. With this information, its initial 3D positions can be estimated by the methods described in section 2.2. BA then makes use of every frame in which a given landmark is detected to reduce 3D location uncertainty. Surface topographic information is extracted from 3D data on the image patch and recorded in the landmark catalog. Figure 7 shows 372 landmark patches (red faces) and corresponding surface normals (arrows). We find that some landmarks are in flatter regions than others. This information is used to select those landmarks best suited for descriptor based recognition in later imagery. Because BA cannot recover the scale directly from image data, we used the ground truth poses of the first two images in the sequence to constrain scale and propagate this to the rest of the sequence. In a real mission, this scale data would be provided by the navigation filter.

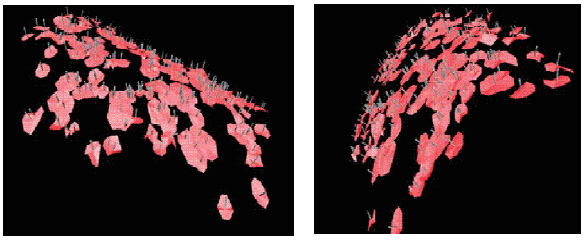

Figure 7: Two views of the landmark topographies and surface normal (arrows in the center of each landmark) of 372 SIFT landmarks.

Figure 8 shows the BA derived pose of the spacecraft for the first 40 images of the sequence. The RMS error in X, Y, and Z direction are 0.337, 0.373 and 0.391 meters and the attitude errors in the X, Y, Z axes are 0.00173, 0.00116, and 0.00130 radian respectively. Since there is no ground truth data on landmarks, which are purely image derived, we estimate landmark localization accuracy through Monte-Carlo simulation assuming the above errors for spacecraft pose accuracy. Given the nominal trajectory shown in Figure 8 and visibility of any landmark in 20 frames, we find through Monte-Carlo simulation that the triangulation error of the 3D landmark positions is on the order of 0.5 m lateral (X and Y) and 0.7 m in depth (Z). This data can be incorporated directly into the navigation filter and can also be used by the pose estimation algorithm in section 2.4 to derive an error covariance on the vision-based pose estimate

It follows that BA can achieve sub-meter accuracy in both spacecraft pose determination and landmark 3D location estimation from an altitude of 300 meters.

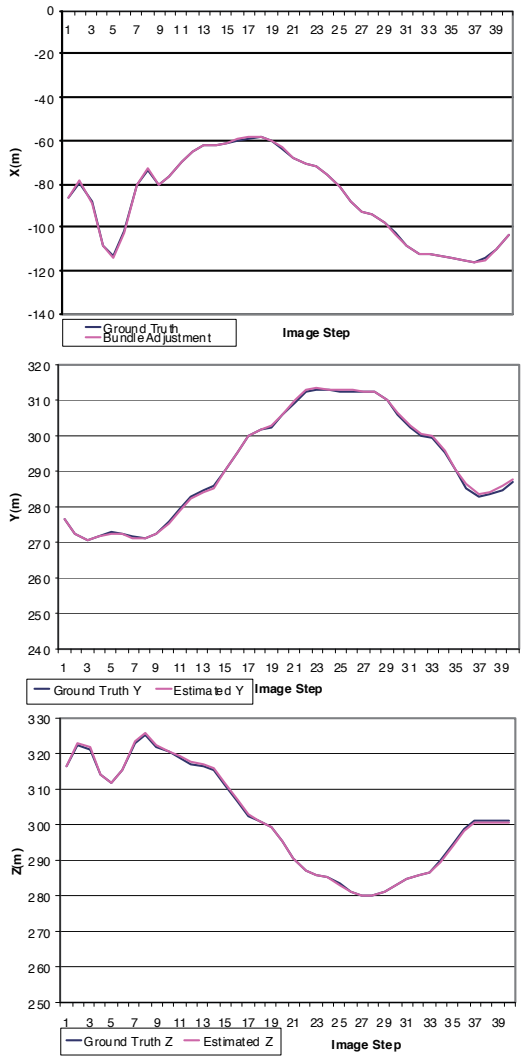

Figure 8: Spacecraft pose estimates on synthetic Itokawa data from BA (red) vs. ground truth (blue).

We overlaid the extracted landmarks and their local topographical properties directly on the Itokawa shape model in Figure 9. Visual inspection tells us the topographies of the landmark match the shape model well, lending further credence to the approach.

We validate spacecraft localization using landmark matching to the catalog during the second descent. Unlike the limited catalog of the "twice around" study, we use the full catalog developed during the first descent toward the Itokawa model in the "twice down" study. Recall that the spacecraft trajectory during this second descent is similar but not identical to the first, hence the rendered imagery differs. While the catalog was generated only during the hover phases, matching and localization was performed throughout the trajectory,

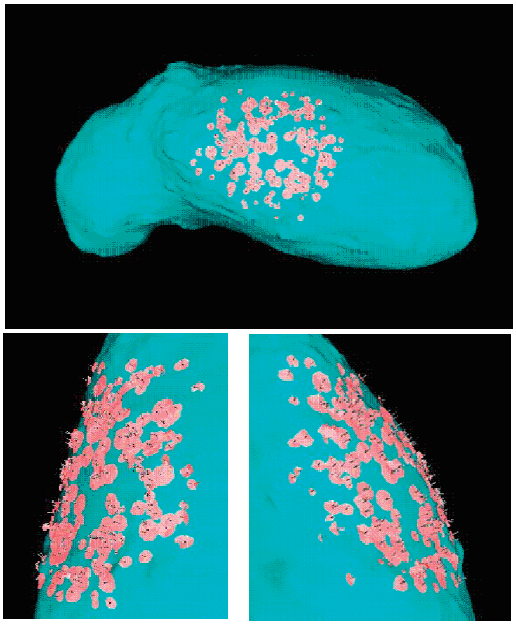

Figure 9: Extracted landmarks overlaid on Itokawa shape model indicating local surface properties.

Results for localization accuracy are shown in Figure 10. RMS errors at 300m were (ΔX, ΔY, ΔZ) = (3.56m, 1.16m, 2.35m) while at 150m, they were (ΔX, ΔY, ΔZ) = (0.43m, 0.38m, 0.91m). While the landmark catalog is only computed at 300m and 150m, not during the intermediate descent, localization errors are well constrained throughout. The missing data in the middle of the descent is due to an image rendering error rather than to failure of landmark matching.

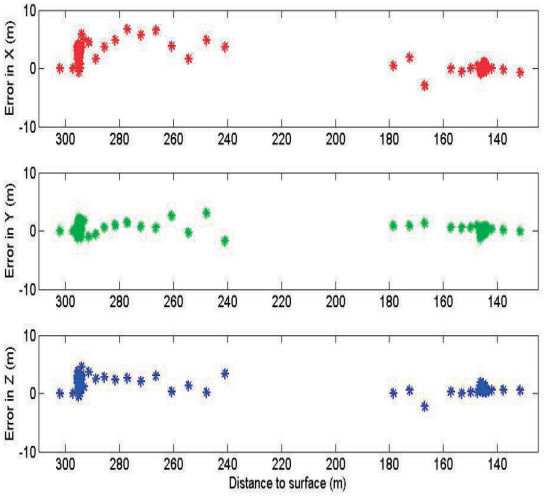

Figure 10: Spacecraft pose estimation errors from twice down study

4 Conclusion

We assumed a plausible two-stage scenario for a small body mission, with one or more stand-off survey phases and a final proximity phase. Given this scenario, we have shown that image based technologies provide a practical solution for spacecraft localization during the proximity phase. In conjunction with a navigation filter, our approach provides enough localization accuracy for targeting of specific sites on the body for sample return or other scientific exploration.

Our approach was to detect distinctive landmarks during the survey phase whose identities could be efficient encoded with a unique descriptor based on image data. These landmarks are recorded in a catalog along with derived 3D locations and additional information to enhance recognition in the proximity phase. Once detected and recognized in the proximity phase, we use image-based pose estimation from 2D-3D correspondence to estimate the spacecraft position and attitude relative to the body. This solution can be used in its own right or as backup to a navigation filter that directly incorporates the 2D-3D correspondences.

We showed through a series of experiments on synthetic and real data that our approach produces highly accurate landmark localization as well as good spacecraft pose, even in the absence of a navigation filter.

5 Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Funding for this work was provided through the internal Research and Technology Development program.

References

[1] Ansar, A., Small Body GN&C Research Report: Feature Recognition Algorithms, *JPL Internal Document D-30282* in Small Body Guidance Navigation and Control FY 2004 RTD Annual Report, *JPL Internal Document D-30714*, pp. 151-171, 2004.

[2] A. Ansar, C. Castano, and L. Matthies, "Enhanced Real-time Stereo Using Bilateral Filtering, *Proc. 2nd Intel. Symp. on 3D Data Processing, Visualization & Transmission (3DPVT)*, Thessaloniki, Greece, 2004.

[3] Ansar, A. and K. Daniilidis, "Linear Pose Estimation from Points and Lines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:578-589, 2003.

[4] Bayard, D.S. and P.B. Brugarolas, "An On-Board Vision-Based Spacecraft Estimation Algorithm for Small Body Exploration," *IEEE Transactions on Aerospace and Electronic Systems.,* Vol. 44, No. 1, pp. 243-260, 2008.

[5] Cheng, Y., Small Body GN&C Research Report: G-CAT – An automated landmark catalog generation tool for renezvous-based exploration, *JPL Internal Document D-36118*, 2006.

[6] Fischler, M. and R. Bolles, "Random sample consensus: A paradigm for model fitting with application to image analysis and autmated cartography," *Communications of the ACM,* vol. 24, pp. 381–395, 1981.

[7] Lowe, D.G., "Distinctive Image Features from Scale-invariant Keypoints,", International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[8] Lu, C-P., G. Hager, and E. Mjolsness, "Fast and Globally Convergent Pose Estimation from Video Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 22:610-622, 2000

[9] R. Hartley and A. Zisserman, Multiple view geometry in computer vision, Cambridge University Press, second edition, 2003.

[10] E. Trucco and A. Verri, Introductory techniques for 3-D computer vision, Prentice Hall, 1998.

[11] Manolis Lourakis and Antonis Argyros, "The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg-Marquardt algorithm", FORTH-ICS / TR-340, 2004.

[12] Takeo Kanade, M. Okutomi, and T. Nakahara, "A multiple-baseline Stereo Method," *Proceedings of DARPA Image Understanding Workshop*, Jan. 27-29 Page 409-426, 1992.