DEEP FEATURE DETECTION AND DESCRIPTION FOR SMALL BODY RELATIVE NAVIGATION

Travis Driver[1*], Katherine Skinner[2], Mehregan Dor[1], and Panagiotis Tsiotras[1]; [1]Georgia Institute of Technology, Atlanta, GA. [2]University of Michigan, Ann Arbor, MI. [*]travisdriver@gatech.edu

Abstract. *Missions to small bodies rely heavily on optical feature tracking for characterization and relative navigation of the target body. While deep learning has led to great advancements in computer vision, training and validating data-driven models for space applications is challenging due to the limited availability of relevant annotated data. Therefore, this paper introduces a large-scale dataset comprised of over 115,000 annotated, real images of 16 different small bodies captured during past and ongoing missions. First, we perform an exhaustive evaluation of both handcrafted and data-driven methods for feature detection and description on small body imagery. Finally, we train a deep feature detection and description network using our data and demonstrate increased performance.*

Introduction. There has been an increasing interest in missions to small bodies (e.g., asteroids, comets) due to their great scientific value.[1] Feature tracking is an integral component of small body shape reconstruction and relative navigation methodologies. However, the current state-of-the-practice in small body relative navigation relies on a manual approach where human operators extract salient surface features from images acquired during an extended characterization phase to estimate a collection of digital terrain maps (DTMs), local topography and albedo maps. DTM construction typically involves extensive human-in-the-loop verification and carefully designed image acquisition plans to achieve optimal results.[2,3] This lack of autonomy in mission procedures limits mission capabilities and increases operational costs.

While automated feature tracking methods have been investigated for autonomous small body relative navigation,[4] researchers have had trouble capitalizing on recent advances in deep learning due to the unavailability of relevant, annotated data.[5,6] To the best of our knowledge, there exists no large-scale, annotated dataset comprised entirely of *real* small body images. Indeed, previous work has either relied entirely on simulated data[7] or uncomprehensive (i.e., <150 images or restricted to a single body) sets of real imagery.[5,8] Moreover, operation in space presents unique environmental (e.g., dynamic hard lighting) and operational (e.g., significant scale and perspective changes) challenges that are likely not adequately captured in datasets based on terrestrial imagery.

This paper presents a large-scale dataset comprised of 115,970 densely annotated, real images of 16 different small bodies from legacy and ongoing deep space missions to facilitate the study of deep learning for autonomous navigation in the vicinity of a small body. The contributions of this paper are as follows: (i) we present a *first-of-a-kind* dataset for vision-based tasks in the vicinity of a small body; (ii) we develop a novel benchmarking suite and evaluate both handcrafted and data-driven feature detection and description methods on real small body imagery; (iii) we train a deep feature detection and description network on small body imagery and demonstrate increased performance on multiple benchmarks.

Data Generation. We leverage publicly available images and ancillary data (i.e., camera pose, camera calibration, shape models) from both legacy and active small body missions provided through NASA's Planetary Data System (PDS).[9] Shape models (i.e., watertight, triangular surface meshes) are developed as part of the relative navigation pipeline of small body missions, typically constructed using stereophotoclinometry.[10] We leverage high-resolution shape models to generate dense depth maps for each image, which are computed by backward-projecting rays at each pixel in the image and recording the intersection depth between the ray and shape model. We also estimate a mask of the occluded regions of the imaged surface for training (see Figure 1).

Evaluation. Performance is evaluated on a per image pair basis using the metrics *precision* = # correct matches / # putative matches, *recall* = # correct matches / # ground truth matches, and *accuracy* = # correct matches & nonmatches / # features. Each method is limited to extract 5,000 features, and putative matches are computed using mutual nearest neighbors. Matches are verified by projecting each keypoint in the first image into the second using the ground truth pose, calibration, and depth map, and matches are verified by checking that the projected coordinates are within 5 pixels of its match. Ground truth matches are registered if there exists a keypoint within 5 pixels of the projected image coordinate. We classify correct nonmatches as keypoints which were not included in the set of putative or ground truth matches.

Finally, poses are computed from the putative matches by first estimating the essential matrix using the five-point method[11] and RANSAC with an inlier threshold of 1 pixel, followed by SVD of the essential matrix to determine the relative pose. We report the area under the curve (AUC) of the cumulative error curve between the estimated and ground truth poses at thresholds of 5°, 10°, and 20°, where the pose error is the maximum of the angular errors in rotation and translation.[12]

We benchmark both *handcrafted* (i.e., SIFT[13]) and *data-driven* (i.e., SuperPoint,[14] R2D2,[15] and ASLFeat[16]) features. These results are summarized in Table 1. SIFT demonstrates competitive performance but suffers when applied to datasets with harsher illumination. SuperPoint achieves high recall but low precision and generally under performs with respect to all other methods. ASLFeat con-

Table 1. Feature performance with respect to precision (P), recall (R), accuracy (A), and pose AUC.

							AUC		
Dataset	# Images	Feature	# Matches	P	R	A	@5°	@10°	@20°
Cassini @ Epimetheus	133	SIFT	204	**32.5**	**36.6**	54.7	2.7	**9.5**	**15.0**
		SuperPoint	396	13.6	26.1	59.2	2.6	7.5	12.8
		R2D2	**423**	25.3	26.1	**77.1**	**2.9**	9.1	14.7
		ASLFeat	386	27.4	29.0	74.7	2.7	8.2	13.7
Cassini @ Mimas	307	SIFT	340	14.3	15.1	41.1	**0.2**	**0.2**	**0.4**
		SuperPoint	121	8.6	10.4	50.5	0.0	0.0	0.0
		R2D2	209	13.8	8.8	**75.5**	0.1	0.1	0.1
		ASLFeat	**372**	**21.8**	**15.7**	65.3	**0.2**	**0.2**	0.3
Dawn @ Ceres	3624	SIFT	**1656**	42.3	72.2	69.4	**28.8**	**44.3**	**56.6**
		SuperPoint	442	42.9	**75.7**	70.1	13.1	28.3	43.5
		R2D2	954	**50.0**	52.8	**85.8**	8.9	20.0	32.4
		ASLFeat	1535	48.4	67.8	80.2	12.9	27.1	42.4
Dawn @ Vesta	2006	SIFT	1350	37.1	52.3	64.0	**17.9**	28.7	38.8
		SuperPoint	506	38.7	55.0	65.8	11.3	21.3	32.7
		R2D2	926	55.9	46.7	**86.9**	11.4	22.3	34.1
		ASLFeat	**1524**	**59.0**	**66.1**	84.3	17.5	**31.9**	**46.0**
Hayabusa @ Itokawa	603	SIFT	217	4.8	5.0	35.8	1.9	3.3	4.8
		SuperPoint	79	7.3	**12.7**	42.3	1.7	3.1	5.4
		R2D2	**339**	10.7	9.4	**67.0**	**2.6**	**4.6**	**8.0**
		ASLFeat	338	**13.5**	11.3	47.5	2.2	4.2	7.6
OSIRIS-REx @ Bennu	1789	SIFT	1317	13.7	15.3	55.2	5.6	8.8	11.8
		SuperPoint	747	18.1	20.3	55.4	3.8	7.3	11.1
		R2D2	502	29.3	18.3	**84.7**	4.2	8.6	13.8
		ASLFeat	**1378**	**33.1**	**30.9**	68.7	**8.0**	**14.4**	**20.9**
Rosetta @ 67P	3039	SIFT	**1168**	15.7	16.6	44.7	2.4	4.8	7.7
		SuperPoint	485	17.6	20.7	49.9	1.6	3.6	6.4
		R2D2	634	20.2	16.5	**79.3**	1.9	3.9	7.1
		ASLFeat	1147	**25.0**	**24.0**	62.8	**3.4**	**6.4**	**10.6**
Rosetta @ Lutetia	40	SIFT	283	23.7	31.7	46.6	5.9	9.8	15.9
		SuperPoint	381	26.7	30.7	55.5	4.2	8.0	16.2
		R2D2	588	33.2	25.6	**74.7**	3.1	6.0	13.3
		ASLFeat	**970**	**42.9**	**35.0**	71.9	**6.0**	**12.1**	**23.8**

Table 2. ASLFeat-CVGBEDTRPJMU compared to pretrained model.

						AUC		
Dataset	Feature	# Matches	P	R	A	@5°	@10°	@20°
Cassini @ Epimetheus†	ASLFeat	386	27.4	**29.0**	**74.7**	**2.7**	8.2	13.7
	ASLFeat-CVGBEDTRPJMU	**396**	**28.9**	27.5	74.1	**2.7**	**8.6**	**14.0**
Cassini @ Mimas†	ASLFeat	**372**	21.8	**15.7**	65.3	**0.2**	**0.2**	**0.3**
	ASLFeat-CVGBEDTRPJMU	328	**23.6**	14.9	**67.1**	0.0	0.1	0.2
Dawn @ Ceres	ASLFeat	**1535**	48.4	67.8	80.2	12.9	27.1	42.4
	ASLFeat-CVGBEDTRPJMU	1514	**52.8**	**71.5**	**82.1**	**15.9**	**31.6**	**46.9**
Dawn @ Vesta	ASLFeat	**1524**	59.0	66.1	84.3	**17.5**	31.9	46.0
	ASLFeat-CVGBEDTRPJMU	1412	**70.3**	**69.7**	**87.4**	**17.5**	**33.0**	**48.7**
Hayabusa @ Itokawa†	ASLFeat	338	13.5	**11.3**	47.5	2.2	4.2	7.6
	ASLFeat-CVGBEDTRPJMU	**363**	**15.2**	11.0	**53.7**	**2.9**	**5.0**	**8.8**
OSIRIS-REx @ Bennu	ASLFeat	**1378**	33.1	**30.9**	68.7	**8.0**	**14.4**	**20.9**
	ASLFeat-CVGBEDTRPJMU	858	**34.2**	28.4	**79.5**	6.7	12.6	19.3
Rosetta @ 67P	ASLFeat	**1147**	25.0	**24.0**	62.8	3.4	6.4	10.6
	ASLFeat-CVGBEDTRPJMU	837	**30.4**	23.9	**69.8**	**4.2**	**7.9**	**13.4**
Rosetta @ Lutetia†	ASLFeat	**970**	**42.9**	**35.0**	71.9	6.0	12.1	**23.8**
	ASLFeat-CVGBEDTRPJMU	778	41.3	31.1	**76.3**	**8.4**	**13.2**	22.3

† No images of this body were included in the training set

sistently ranks among the top performing methods with respect to all datasets. Therefore, we selected ASLFeat for end-to-end training.

Learning Features from Small Body Imagery. We train ASLFeat using a procedure similar to the original implementation[16] with an approximate 90/10 train/test split. We withhold data corresponding to 4 different small bodies with variable surface characteristics from training, i.e., Epimetheus, Mimas, 25143 Itokawa, and 21 Lutetia. This will test the network's ability to reliably compute features upon arrival at a previously unexplored body. The network is also tested on held-out images of small bodies it observed during training. The ASLFeat model trained on small body imagery, i.e., ASLFeat-CVGBEDTRPJMU, is compared against the pretrained model in Table 2. ASLFeat-CVGBEDTRPJMU consistently outperforms the pretrained model with respect to our metrics. Importantly, our model achieves higher matching precision and pose AUC on almost all novel testing instances.

Acknowledgements. This work was supported by a NASA Space Technology Graduate Research Opportunity.

References.

[1] M. A. Barucci, E. Dotto, and A. C. Levasseur-Regourd, "Space missions to small bodies: asteroids and cometary nuclei," *Astronomy and Astrophysics Review*, vol. 19, no. 48, pp. 1–29, 2011.

[2] O. S. Barnouin et al., "Digital terrain mapping by the OSIRIS-REx mission," *Planetary and Space Science*, vol. 180, p. 104764, 2020.

[3] E. Palmer et al., "Practical stereophotoclinometry for modeling shape and topography on planetary missions," *Planetary Science*, vol. 3, no. 102, pp. 1–16, 2022.

[4] B. J. Morrell, J. Villa, and A. Havard, "Automatic feature tracking on small bodies for autonomous approach," in *ASCEND*, pp. 1–15, 2020.

[5] T. Fuchs et al., "Enhanced flyby science with onboard computer vision: Tracking and surface feature detection at small bodies," *ESS*, vol. 2, no. 10, pp. 417–434, 2015.

[6] J. Song, D. Rondao, and N. Aouf, "Deep learning-based spacecraft relative navigation methods: A survey," *Acta Astronautica*, vol. 191, pp. 22–40, 2022.

[7] M. Pugliatti, M. Maestrini, P. Di Lizia, and F. Topputo, "On-board small-body semantic segmentation based on morphological features with U-Net," in *AAS/AIAA Space Flight Mechanics Meeting*, pp. 1–20, 2021.

[8] H. Lee, H.-L. Choi, D. Jung, and S. Choi, "Deep neural network-based landmark selection method for optical navigation on lunar highlands," *IEEE Access*, vol. 8, pp. 99010–99023, 2020.

[9] "NASA PDS." `https://pds.nasa.gov/`.

[10] R. W. Gaskell et al., "Characterizing and navigating small bodies with imaging data," *Meteoritics & Planetary Science*, vol. 43, no. 6, pp. 1049–1061, 2008.

[11] D. Nistér, "An efficient solution to the five-point relative pose problem," *TPAMI*, vol. 26, no. 6, pp. 756–770, 2004.

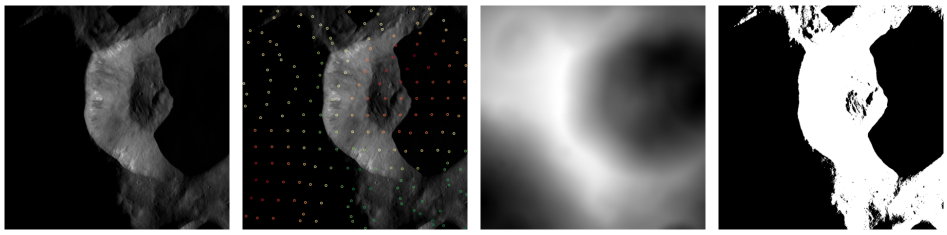
[12] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperGlue: Learning feature matching with graph neural networks," in *CVPR*, pp. 4938–4947, 2020.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[14] D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-supervised interest point detection and description," in *CVPR Workshops*, pp. 337–349, 2018.

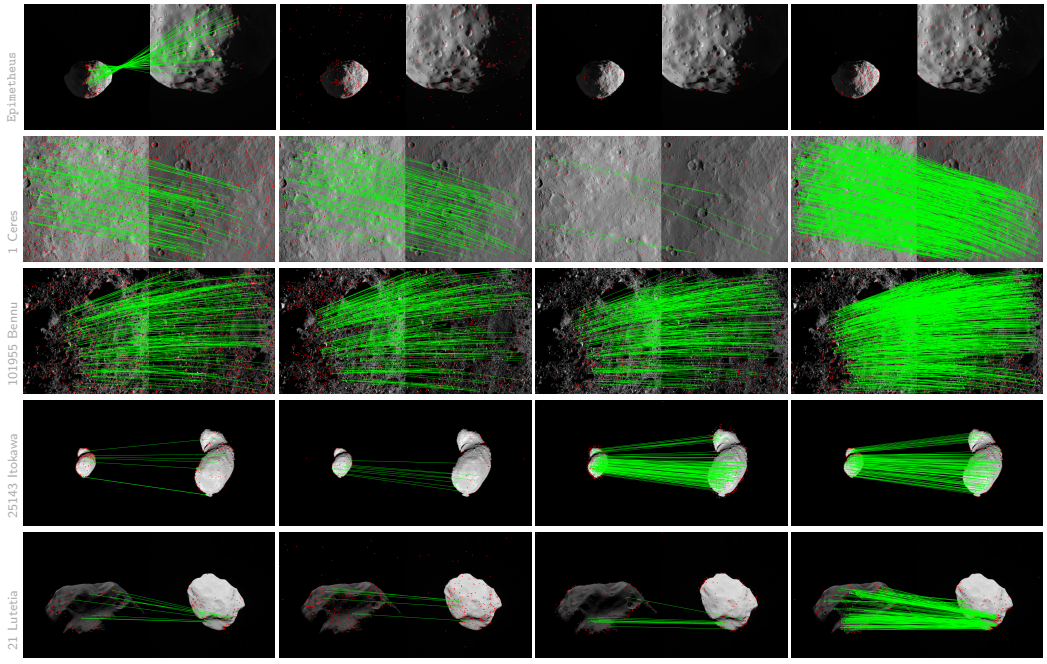
[15] J. Revaud, C. De Souza, M. Humenberger, and P. Weinzaepfel, "R2D2: Reliable and repeatable detector and descriptor," in *NeurIPS*, pp. 1–11, 2019.

[16] Z. Luo et al., "ASLFeat: Learning local features of accurate shape and localization," in *CVPR*, pp. 6589–6598, 2020.



(a) Image *(b) Landmark map* *(c) Depth map* *(d) Mask*

Figure 1. Example data products.



SIFT ***SuperPoint*** ***R2D2*** ***AsLFeat***

Figure 2. Qualitative comparison of matching performance. Correct matches are drawn in green and the keypoints of incorrect matches are drawn in red.

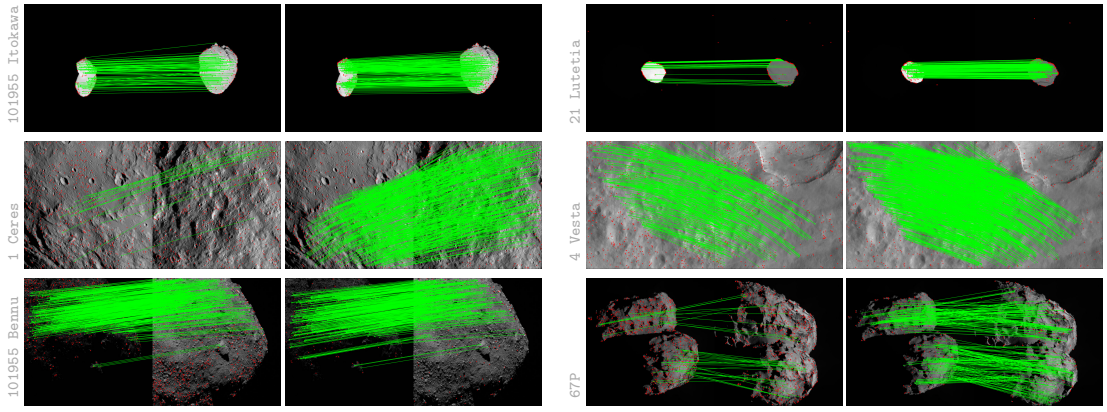


Figure 3. Qualitative comparison between pretrained model (left) and ASLFeat-CVGBEDTRPJMU (right) feature matches. Correct matches are drawn in green, and the keypoints of incorrect matches are drawn in red.