Night-Voyager: Consistent and Efficient Nocturnal Vision-Aided State Estimation in Object Maps*

Tianxiao Gao[1⋆], Mingle Zhao[1], Chengzhong Xu[1], and Hui Kong[1]
[1]University of Macau, ⋆Corresponding Author
Email: {ga0.tianxiao; zhao.mingle}@connect.um.edu.mo; {czxu; huikong}@um.edu.mo

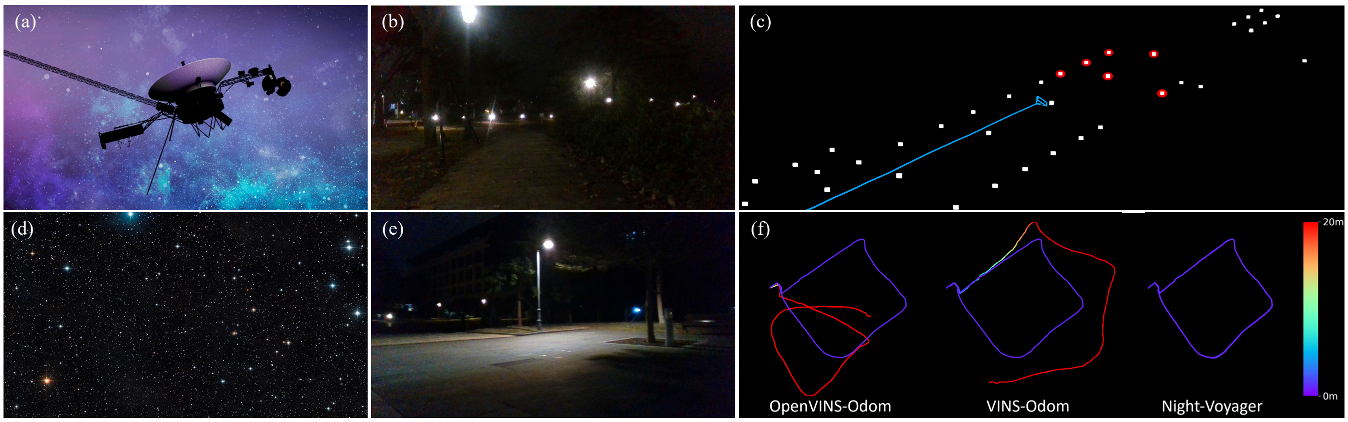

Fig. 1: Inspired by the interstellar navigation of spacecrafts, which exploits prior ephemeris data, star coordinates, and onboard star images to achieve localization [1]–[3], we propose Night-Voyager, a framework utilizing prior object-level streetlight maps for nocturnal vision-aided state estimation. (a) and (d) depict the NASA's Voyager mission [4] and a typical image of star systems in deep space [5] (image credit: NASA/JPL-Caltech and ESO/Digitized Sky Survey). (b) and (e) are typical images captured in nocturnal scenes. (c) displays the online localization (the blue curve) of Night-Voyager within the streetlight map (white boxes) and the matches (red spheres). (f) depicts the trajectories estimated by the odometer-aided OpenVINS [6], the odometer-aided VINS-Mono [7], [8], and Night-Voyager [9], respectively. The color bar indicates the trajectory error scale with respect to the ground truth (purple curves).

***Abstract*—Accurate and robust state estimation at nighttime is essential for autonomous robotic navigation to achieve nocturnal or round-the-clock tasks. An intuitive question arises: Can low-cost standard cameras be exploited for nocturnal state estimation? Regrettably, most existing visual methods may fail under adverse illumination conditions, even with active lighting or image enhancement. A pivotal insight, however, is that streetlights in most urban scenarios act as stable and salient prior visual cues at night, reminiscent of stars in deep space aiding spacecraft voyage in interstellar navigation. Inspired by this, we propose Night-Voyager, an object-level nocturnal vision-aided state estimation framework that leverages prior object maps and keypoints for versatile localization. We also find that the primary limitation of conventional visual methods under poor lighting conditions stems from the reliance on pixel-level metrics. In contrast, metric-agnostic, non-pixel-level object detection serves as a bridge between pixel-level and object-level spaces, enabling effective propagation and utilization of object map information within the system. Through comprehensive experiments in both simulation and diverse real-world scenarios, Night-Voyager showcases its efficacy, robustness, and efficiency, filling a critical gap in nocturnal vision-aided state estimation.**

I. Visual Methodology in Nocturnal Scenes

1) Conventional Vision-aided State Estimation: Although these methods [6]–[8], [10]–[14] achieve great performance in well-lit and textured environments, they are prone to failures in nocturnal environments, as shown in Fig. 1. The unfavorable, unstable, and inconsistent illumination conditions lead to limited visual information and erroneous data association.

2) Active Lighting for Illumination Enhancement: Intuitively, a practical solution is to utilize active headlights in low-light scenarios. However, as shown in Fig. 2, active lighting can exacerbate the inconsistent and imbalanced variance of illumination [15]–[17]. Moreover, illuminated dust particles [18], including high-reflectivity objects [16] are sparkling due to backscattering [19], leading to chaotic data association.

3) Image Quality Enhancement: There are numerous classical image processing methods [20]–[22] and deep learning-based methods [23]–[28] to enhance low-light images. Nevertheless, the generalization and real-time performance of these methods fail to meet the requirements of robotic applications. Furthermore, image enhancement is also challenging in addressing the inconsistency issue between images.

4) Data Association Enhancement: Extensive works are proposed to achieve consistent feature matching [29]–[38]. However, applying these methods in low-light environments remains challenging. In addition, the issues of generalization, parameter tuning, and real-time performance also represent bottlenecks for online robotic applications.

II. Crux and Key Insight

The underlying problem of nocturnal visual state estimation stems from two primary bottleneck factors: *insufficiency* and

*__Publication information__: Gao, Tianxiao, Mingle Zhao, Chengzhong Xu, and Hui Kong. "Night-Voyager: Consistent and Efficient Nocturnal Vision-Aided State Estimation in Object Maps." *IEEE Transactions on Robotics* (2025).
DOI: https://doi.org/10.1109/TRO.2025.3548540.
Code and files: https://github.com/IMRL/Night-Voyager.
Supplementary video: https://youtu.be/x7XtC_ALz80.

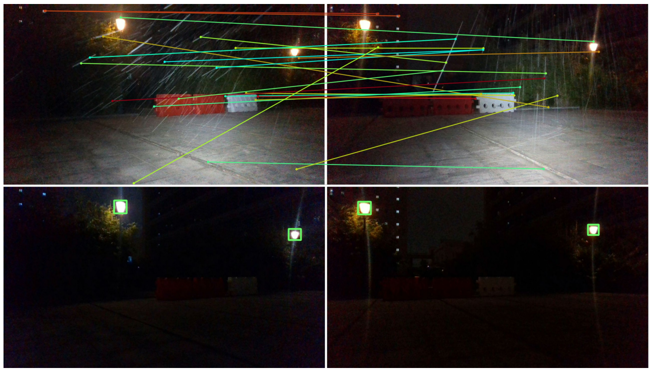

Fig. 2: Top row: active lighting results in a significant number of erroneous feature matches (colored dots and lines). Bottom row: object-level detection of streetlights (green detection boxes) remains extraordinarily robust even in completely dark nighttime scenarios.

inconsistency. The insufficiency indicates the lack of visual features in low-light environments. Even when features are present, they are often transient and typically originate from dynamic objects. The two key bottleneck factors point to a crux: conventional visual methods heavily rely on *pixel-level* feature extraction and data association. The following key insights motivate us to fundamentally solve the problem:

Key Insight 1 *(Pixel-level and Pixel-metric Methodologies). Low-light conditions lead to insufficiency and inconsistency issues stemming from pixel-level and pixel-metric methodologies, while metric-agnostic and object-level methods are immune to these two bottlenecks.*

Key Insight 2 *(Prior is All You Need). This work reinforces a seminal insight that remains equally essential for both model-based and data-driven methodologies: Prior is all you need–whether from sensors, physical models, or learned models. Essentially, prior information implicitly provides reliable constraints, effectively mitigating the impact of significant model or sensor noise, particularly in challenging environments.*

Key Insight 3 *(Consistent and Efficient Estimator). A consistent and efficient estimator is essential for measurements with substantial uncertainties in challenging environments.*

III. Night-Voyager

In this work, we propose Night-Voyager, a hybrid object- and pixel-level vision-aided state estimation framework, as shown in Fig. 3. We briefly introduce the contained modules as follows, and the work details can be referred to [9].

1) Multi-Sensor Fusion Module (MSF): The MSF module that fuses sensor measurements is independent of other modules, enabling the system to achieve all-day state estimation.

2) Initialization: To determine the initial pose in the map without the aid of Global Navigation Satellite Systems (GNSS), the initialization module divides the map-based global localization problem into a series of *Perspective-Three-Point* (P3P) problems [39]. With the two-level filtering method, the optimal solution can be determined quickly and accurately.

3) Map-Based Localization: A two-stage cross-modal data association approach for streetlighs in this module ensures accuracy and robustness of the detection-map matching process, providing reliable object-level observations for state update.

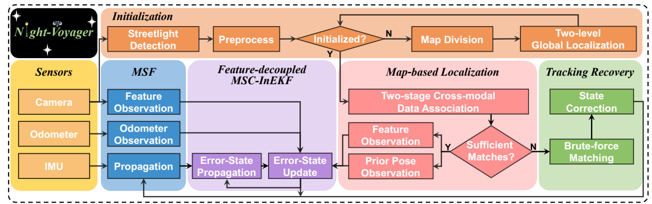

Fig. 3: System overview of the proposed Night-Voyager.

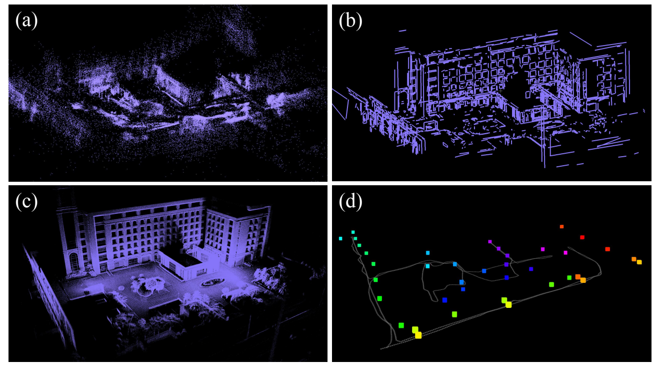

Fig. 4: Different map representations for vision-aided state estimation. (a) Visual feature point map. (b) Line map. (c) Point cloud map. (d) Streetlight map where the colored blocks represent streetlight instances and the gray ones are prior poses from the mapping process. Compared with (a), (b), and (c), (d) is a highly lightweight map containing both geometric and semantic object-level information.

4) Tracking Recovery: For improving robustness, this module is designed to correct the state estimate by using the brute-force method when streetlight matches are absent for a while.

5) Feature-Decoupled MSC-InEKF: The proposed filter endows Night-Voyager with a consistent and efficient state estimator. The Lie group-based filter with multi-state constraint makes it particularly robust in nocturnal state estimation, even in scenarios with high observation noise, prolonged absent observations, and considerable initialization errors.

We perform comprehensive experiments in both simulation and real-world scenarios, as detailed in [9]. Fig. 4 shows different maps commonly used in visual localization methods [33], [40]–[42]. However, all of them achieve poor performance in nighttime cases. The essential reason lies in the reliance on the pixel-level features and metrics. The streetlight map, which provides object-level features, empowers Night-Voyager with accurate and robust state estimation in low-light environments.

IV. Discussion and Conclusion

In this work, we propose Night-Voyager, an object-level vision-aided state estimation framework to fundamentally resolve the *insufficiency* and *inconsistency* bottlenecks in low-light visual tasks. The genesis of the bottlenecks is the reliance on pixel-level and pixel-metric methodologies. In contrast, object-level avenues are immune to inconsistency and transience, leading to a fundamental solution for nocturnal visual problems. Significantly, this work also reinforces the importance of prior information for both model-based and data-driven methodologies.

A promising direction for future work is the integration of semantic cues or object tracking to impose temporal consistency without dependence on prior maps, advancing towards a unified all-day navigation framework.

REFERENCES

[1] R. JACOBSON, S. SYNNOTT, and J. DIXON, "Navigation accuracy analysis for the mariner-jupiter-saturn 1977 mission," in *Astrodynamics Conference*, 1976, p. 833.

[2] C. Kohlhase and P. A. Penzo, "Voyager mission description," *Space science reviews*, vol. 21, no. 2, pp. 77–101, 1977.

[3] NASA Jet Propulsion Laboratory, "First images from nasa's europa clipper," https://www.jpl.nasa.gov/images/pia26561-first-images-from-nasas-europa-clipper/.

[4] NASA Science Editorial Team, "Nasa turns off 2 voyager science instruments to extend mission," https://science.nasa.gov/blogs/voyager/2025/03/05/nasa-turns-off-2-voyager-science-instruments-to-extend-mission/.

[5] Space.com Staff, "Around a star system — space wallpaper," https://www.space.com/27600-around-a-star-system-space-wallpaper.html.

[6] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "Openvins: A research platform for visual-inertial estimation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 4666–4672.

[7] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[8] J. Liu, W. Gao, and Z. Hu, "Visual-inertial odometry tightly coupled with wheel encoder adopting robust initialization and online extrinsic calibration," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 5391–5397.

[9] T. Gao, M. Zhao, C. Xu, and H. Kong, "Night-voyager: Consistent and efficient nocturnal vision-aided state estimation in object maps," *IEEE Transactions on Robotics*, 2025.

[10] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "Svo: Semidirect visual odometry for monocular and multicamera systems," *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 249–265, 2016.

[11] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 3, pp. 611–625, 2017.

[12] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE transactions on robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[13] M. Bloesch, M. Burri, S. Omari, M. Hutter, and R. Siegwart, "Iterated extended kalman filter based visual-inertial odometry using direct photometric feedback," *The International Journal of Robotics Research*, vol. 36, no. 10, pp. 1053–1072, 2017.

[14] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[15] J. Kim, Y. Cho, and A. Kim, "Proactive camera attribute control using bayesian optimization for illumination-resilient visual navigation," *IEEE Transactions on Robotics*, vol. 36, no. 4, pp. 1256–1271, 2020.

[16] J. Kim, M.-H. Jeon, Y. Cho, and A. Kim, "Dark synthetic vision: Lightweight active vision to navigate in the dark," *IEEE Robotics and Automation Letters*, vol. 6, no. 1, pp. 143–150, 2020.

[17] Y. Lin, S. Wang, Y. Jiang, and B. Han, "Breaking of brightness consistency in optical flow with a lightweight cnn network," *IEEE Robotics and Automation Letters*, 2024.

[18] C. F. Bohren and D. R. Huffman, *Absorption and scattering of light by small particles*. John Wiley & Sons, 2008.

[19] T. Grimm and M. Grimm, *The Basic Book of Photography*, ser. Basic Book of Photography. Plume, 1997. [Online]. Available: https://books.google.com/books?id=xIMrAQAAMAAJ

[20] A. Yamasaki, H. Takauji, S. Kaneko, T. Kanade, and H. Ohki, "Denighting: Enhancement of nighttime images for a surveillance camera," in *2008 19th International Conference on Pattern Recognition*. IEEE, 2008, pp. 1–4.

[21] X. Dong, Y. Pang, and J. Wen, "Fast efficient algorithm for enhancement of low lighting video," in *ACM SIGGRAPH 2010 posters*, 2010, pp. 1–1.

[22] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides, "Low light image enhancement via sparse representations," in *Image Analysis and Recognition: 11th International Conference, ICIAR 2014, Vilamoura, Portugal, October 22-24, 2014, Proceedings, Part I 11*. Springer, 2014, pp. 84–93.

[23] K. G. Lore, A. Akintayo, and S. Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.

[24] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *IEEE transactions on image processing*, vol. 30, pp. 2340–2349, 2021.

[25] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5637–5646.

[26] S. Yang, M. Ding, Y. Wu, Z. Li, and J. Zhang, "Implicit neural representation for cooperative low-light image enhancement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 12 918–12 927.

[27] Y. Wu, C. Pan, G. Wang, Y. Yang, J. Wei, C. Li, and H. T. Shen, "Learning semantic-aware knowledge guidance for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1662–1671.

[28] W. Wang, H. Yang, J. Fu, and J. Liu, "Zero-reference low-light enhancement via physical quadruple priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 26 057–26 066.

[29] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91–110, 2004.

[30] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9*. Springer, 2006, pp. 430–443.

[31] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *2011 International conference on computer vision*. Ieee, 2011, pp. 2564–2571.

[32] S. Leutenegger, M. Chli, and R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *2011 International conference on computer vision*. Ieee, 2011, pp. 2548–2555.

[33] P.-E. Sarlin, C. Cadena, R. Siegwart, and M. Dymczyk, "From coarse to fine: Robust hierarchical localization at large scale," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 12 716–12 725.

[34] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 224–236.

[35] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superglue: Learning feature matching with graph neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 4938–4947.

[36] J. Sun, Z. Shen, Y. Wang, H. Bao, and X. Zhou, "Loftr: Detector-free local feature matching with transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 8922–8931.

[37] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *International Journal of Computer Vision*, vol. 129, no. 1, pp. 23–79, 2021.

[38] X. Jiang, J. Ma, G. Xiao, Z. Shao, and X. Guo, "A review of multimodal image matching: Methods and applications," *Information Fusion*, vol. 73, pp. 22–71, 2021.

[39] Y. Ding, J. Yang, V. Larsson, C. Olsson, and K. Åström, "Revisiting the p3p problem," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4872–4880.

[40] H. Yu, W. Zhen, W. Yang, J. Zhang, and S. Scherer, "Monocular camera localization in prior lidar maps with 2d-3d line correspondences," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 4588–4594.

[41] D. Cattaneo, M. Vaghi, A. L. Ballardini, S. Fontana, D. G. Sorrenti, and W. Burgard, "Cmrnet: Camera to lidar-map registration," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Oct 2019, pp. 1283–1289.

[42] X. Wu, J. Xu, P. Hu, G. Wang, and H. Wang, "Lhmap-loc: Cross-modal monocular localization using lidar point cloud heat map," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 8500–8506.