SCIENCE CHINA
Information Sciences

• LETTER •

February 2026, Vol. 69, Iss. 2, 129201:1–129201:2
https://doi.org/10.1007/s11432-024-4557-1

Outdoor full-range geolocator with multi-vision fusion and self-supervised truncation filtering

Wenbin LIU[1], Feng GAO[1], Yeyun CAI[1*], Lele ZHANG[1], Yang LIU[1], Hailing FU[1], Chen CHEN[1] & Fang DENG[1,2]

[1]*School of Automation, Beijing Institute of Technology, Beijing 100081, China*
[2]*Beijing Institute of Technology Chongqing Innovation Center, Chongqing 401120, China*

Received 12 October 2024/Revised 4 March 2025/Accepted 14 August 2025/Published online 4 January 2026

Citation Liu W B, Gao F, Cai Y Y, et al. Outdoor full-range geolocator with multi-vision fusion and self-supervised truncation filtering. Sci China Inf Sci, 2026, 69(2): 129201, https://doi.org/10.1007/s11432-024-4557-1

Accurate target localization plays a crucial role in robotics, autonomous navigation, and surveillance applications. Compared to lidar, cameras are more affordable, easier to install, and provide richer semantic information, leading to camera-based target localization methods receiving widespread attention. However, despite their widespread adoption and extensive validation on benchmark datasets, camera-based localization methods still face significant challenges in real-world outdoor applications (for a detailed literature review, please refer to Appendix A). Specifically, two main challenges need to be resolved.

First, cameras with a single focal length cannot maintain accuracy across the full range. Targets that are too close suffer from truncation, while those that are too far experience resolution limitations. Multi-view stereo vision can mitigate this issue by providing additional depth cues, but most existing methods rely on cameras with similar focal lengths [1]. As a result, these approaches enhance depth perception within a fixed working area (WA) but fail to expand it (for a detailed explanation of WA, please refer to Appendix A).

Second, learning-based methods are costly to deploy in real-world applications. Most current public datasets are collected with cameras of similar focal lengths, which often do not match those used in practical scenarios. It is nearly impossible to recollect sufficient data for deployment due to the high cost of 3D annotation. Recent studies have attempted to enhance model generalization across different focal lengths, but they still struggle with images taken at focal lengths outside the datasets range [2] or require fine-tuning on existing 3D datasets for accurate depth estimation [3,4].

To address these challenges, we propose the full-range multi-view geolocator (FMVG), which integrates multiple cameras with different focal lengths to extend the working area while ensuring high localization accuracy. FMVG employs a blind-spot-free WA expansion strategy to unify short- and long-range capabilities seamlessly. Based on our previous work [5], we also introduce a self-supervised truncation detection module to improve localization reliability and enhance generalization across focal lengths, reducing dependence on extensive 3D annotations. Furthermore, a 2D projection association method and maximum likelihood estimation are utilized to fuse localization results from multiple cameras, which are subsequently transformed into absolute coordinates via GPS and visualized on an electronic map for intuitive representation. These components collectively strengthen the robustness and generalization ability of our system while reducing the need for 3D annotations.

The main contributions of this paper are as follows.

(1) We propose a novel outdoor FMVG capable of accurate geolocalization across the full range without any 3D annotations.

(2) We build a WA model for fixed focal length cameras and propose a WA expansion solution for cameras with different focal lengths that eliminates blind spots.

(3) We design a symmetrical multi-camera array and conduct several groups of experiments to validate FMVG in real outdoor environments.

Methodology. The overview of our proposed FMVG framework is shown in Figure 1(a), comprising three parts: blind-spot- free WA expansion solution, target localization with a truncation filter, and multi-vision fusion. Details of these components are provided in Appendix B.

Blind-spot-free WA expansion solution. The WA of a fixed focal length camera is constrained by projection theory. Given a target with height H_o, focal length f, and image projection size h, the target depth T_z is given by

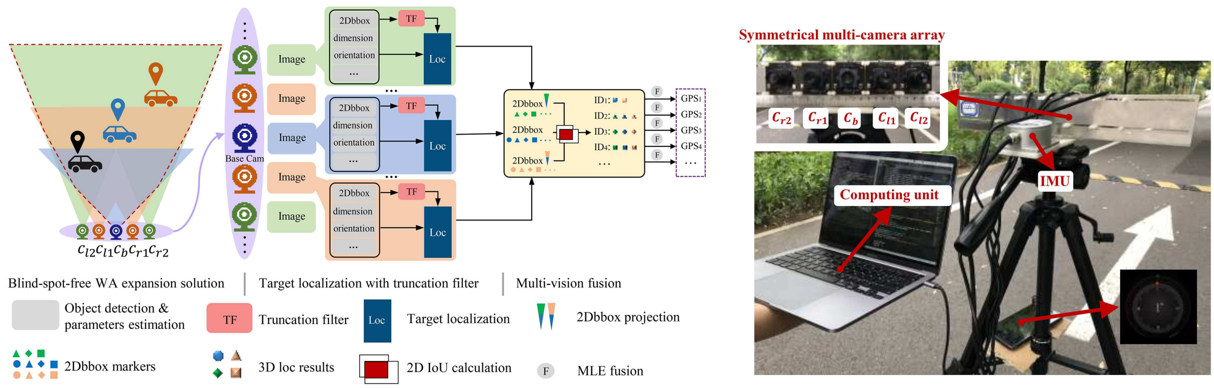
$$T_z = \frac{f H_o}{h}. \tag{1}$$

Based on this relation, we define the upper and lower WA boundaries for a fixed-focal-length camera.

To achieve seamless full-range localization, our method dynamically selects and arranges focal lengths to ensure continuous WA coverage, eliminating blind spots, and enabling precise and robust target localization across varying distances.

Target localization with a truncation filter. Our previous work [5] could recover the target's 3D position from the 2D detection results based on geometric consistency. However, in the multi-focal camera system of the WA expansion solution, truncated targets inevitably lead to unreliable localization results, reducing the effectiveness of fusion-based localization.

* Corresponding author (email: caiyeyun@bit.edu.cn)



(a) **(b)**

Figure 1 (Color online) (a) Overview of FMVG. The blind-spot-free WA expansion solution integrates the WAs of cameras with different focal lengths to maintain accuracy across the full range. In target localization with the truncation filter, each camera performs localization simultaneously and filters out poor results. Finally, multi-vision fusion involves associating and integrating the results to achieve more accurate localization, which is then converted into absolute geographic positions. (b) Multi-vision geolocalization system. The system comprises a symmetrical multi-camera array that contains five cameras, an IMU, an onboard compass, and a computing unit.

To address this issue, we introduce a self-supervised truncation detection module that filters out severely truncated targets before localization, ensuring that only reliable targets contribute to the final estimation. This module assesses target completeness by generating pseudo-truncated images from complete 2D detections using randomly generated masks and assigning truncation labels accordingly. The generated data and labels are then used to train a truncation estimation network, enabling effective identification of severely truncated targets.

Multi-vision fusion. In the WA expansion solution, the target location results of multiple cameras must be accurately associated and fused to improve overall positioning accuracy. To address this, we introduce a 2D projection association method to match results from different cameras on the 2D image plane, establishing correspondences between detected targets and ensuring correct multi-view alignment. We then apply maximum likelihood estimation (MLE) to probabilistically fuse localization results, minimizing uncertainty and improving depth consistency. Finally, the fused localization results are transformed into absolute coordinates using GPS, allowing precise geospatial positioning.

Experimental results. To evaluate the effectiveness of FMVG, we built a physical test platform, as shown in Figure 1(b), and conducted extensive real-world experiments, comparing our approach with state-of-the-art (SOTA) methods. The experiments were designed to assess localization accuracy across varying distances and environmental conditions, demonstrating the robustness and adaptability of our method. We quantitatively evaluate FMVG using key metrics, including average localization error, standard deviation, precision, and recall. Extensive experiments show that FMVG can accurately locate targets in a full range from ∼9 to ∼137 m, with an average accuracy of 0.76 m and an average standard deviation of 0.6 m, outperforming current localization methods.

The high precision and recall of FMVG further demonstrate its stability and reliability in various environments. Ablation studies further validate the contribution of each module, demonstrating that the WA expansion strategy, truncation detection, and multi-view fusion collectively enhance localization performance. Detailed experimental settings and results are provided in Appendix C.

Conclusion. We present a novel outdoor geolocator FMVG capable of accurate geolocalization across the full range without any 3D annotations. The pipeline of FMVG consists of a blind-spot-free WA expansion solution, target localization with a truncation filter, and multi-vision fusion. FMVG improves the overall positioning accuracy through data fusion while extending the joint WA to a full range. Through extensive outdoor experiments, FMVG proved its capability to maintain high performance, achieving an average accuracy of 0.76 m and a standard deviation of 0.6 m across a wide ∼9 to ∼137 m range, validating it as a superior alternative to current methods. In the future, we will further study the multi-focal length feature fusion method and verify our method in actual driving.

Acknowledgements This work was supported in part by National Science Foundation for Distinguished Young Scholars of China (Grant No. 62025301), National Natural Science Foundation of China Basic Science Center Program (Grant No. 62088101), and China National Postdoctoral Program for Innovative Talents (Grant No. BX20240466).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Deng F, Zhang L, Gao F, et al. Long-range binocular vision target geolocation using handheld electronic devices in outdoor environment. IEEE Trans Image Process, 2020, 29: 5531–5541
2 Li Z, Xu X, Lim S, et al. Unimode: unified monocular 3D object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024. 16561–16570
3 Yang L, Kang B, Huang Z, et al. Depth anything: unleashing the power of large-scale unlabeled data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024. 10371–10381
4 Yang L, Kang B, Huang Z, et al. Depth anything v2. 2024. ArXiv:2406.09414
5 Gao F, Deng F, Li L, et al. MGG: monocular global geolocation for outdoor long-range targets. IEEE Trans Image Process, 2021, 30: 6349–6363