Fast and accurate obstacle detection based on stereo vision and deep learning

Samira Badrloo [1,2]

[1] Zanjan Regional Electricity Company, Iran - badrloo.samira@email.kntu.ac.ir

[2]Department of Photogrammetry and Remote Sensing, K.N. Toosi University of Technology, Tehran 19697, Iran

Keywords: Stereo images, Deep learning, Mobile robot, Obstacle detection, Objects.

Abstract

The ability to detect and avoid obstacles is essential for ensuring the safe and efficient navigation of mobile robots. With the increasing demand for intelligent autonomous systems, the need for fast and accurate obstacle detection has become a critical area of research. Existing methods for obstacle detection can be broadly classified into three categories: sensor-based, image-based, and hybrid approaches. Among these, vision-based techniques have gained significant attention due to their effectiveness and versatility. These methods can be further divided into monocular and stereo approaches, each offering distinct advantages. In recent years, stereo vision-based methods have emerged as a promising solution for obstacle detection, as they enable the precise estimation of depth and distance to objects, providing valuable information for real-time navigation. However, despite their accuracy, stereo methods are often criticized for their high computational complexity and processing inefficiency, which can limit their practicality in high-speed robotic applications. This study presents a novel hybrid approach that integrates the advantages of both stereo and monocular methods. By leveraging monocular techniques for object detection and utilizing stereo vision for precise depth estimation, our method enhances efficiency while maintaining high accuracy. Instead of computing depth information for individual pixels, the proposed method operates at the object level, calculating the distance for detected objects rather than processing each pixel separately. In addition, comparative analysis with the method developed by Lamini, Fathi et al. 2024 demonstrates that the proposed approach yields more stable and accurate results, further highlighting its effectiveness and reliability in obstacle detection.

1. Introduction

Obstacle detection is a fundamental requirement for ensuring the safe and efficient operation of mobile robots, allowing them to navigate complex, dynamic environments while avoiding collisions. The effectiveness of obstacle detection methods is generally categorized into three primary approaches: image-based, sensor-based, and hybrid techniques (Goodin, Carrillo et al. 2021, Aharchi and Kbir 2022, Shi, Chang et al. 2023). Among these, image-based methods have gained significant attention due to their ability to process rich visual information, including gray levels, points, edges, and regions (Aguilar, Casaliglla et al. 2017, Al-Kaff, García et al. 2017, Badrloo, Varshosaz et al. 2022). By leveraging these visual cues, image-based approaches provide detailed and accurate environmental perception, thereby enhancing a robot’s real-time decision-making and situational awareness.

As mobile robotics and artificial intelligence continue to advance, image-based obstacle detection is emerging as a highly effective and scalable solution for complex navigation challenges. The integration of advanced computer vision techniques and deep learning models has further improved the robustness and adaptability of these systems, enabling them to perform efficiently across diverse, unstructured environments. Consequently, vision-based obstacle detection remains a crucial component in the development of autonomous robotic systems, significantly enhancing their ability to operate in real-world scenarios (Badrloo, Varshosaz et al. 2022).

Vision-based obstacle detection methods can be broadly categorized into two primary types: (a) monocular methods and (b) stereo methods. Monocular algorithms utilize a single camera to detect obstacles (Lee, Ho et al. 2021, Shi, Chang et al. 2023), while stereo techniques employ two cameras to capture depth information.

Monocular methods can be categorized into four distinct groups: appearance-based, motion-based, depth-based, and expansion-based approaches (Badrloo and Varshosaz 2022). Although monocular obstacle detection is a cost-effective and widely applicable solution, it faces substantial challenges in depth estimation and robustness, particularly in complex environments. To mitigate these limitations, hybrid approaches integrating monocular vision with Inertial Measurement Units (IMUs), LiDAR, or stereo cameras have demonstrated significantly improved reliability, enhancing both accuracy and adaptability in real-world scenarios (Al-Kaff, García et al. 2017, Lee, Ho et al. 2021).

Stereo-based obstacle detection offers a more reliable depth estimation compared to monocular methods, making it suitable for autonomous navigation and robotics. They provide a more accurate and comprehensive understanding of the environment (Barry, Florence et al. 2018, Grinberg and Ruf 2021, Sun, Li et al. 2021). However, its effectiveness is limited by computational demands, environmental constraints, and calibration challenges.

Despite their advantages, stereo methods have significant limitations, particularly in computational efficiency, environmental adaptability, and calibration (Vargas, Alsweiss et al. 2021, Zhang, Yang et al. 2023). These methods typically process data at the pixel level, demanding substantial computational resources and introducing delays that slow obstacle detection. Such delays compromise the accuracy of obstacle identification and hinder real-time navigation. To fully harness the potential of stereo methods, addressing these inefficiencies is essential to achieving faster and more precise obstacle detection for mobile robots.

Real-time and accurate obstacle detection using stereo techniques has long posed a significant challenge, primarily due to the high computational complexity involved (Tijmons, de Croon et al. 2017, Bharati, Wu et al. 2018, Wu, Huang et al. 2018, Liu, Li et al. 2021, Badrloo, Varshosaz et al. 2022). To address this issue, researchers have proposed various methods aimed at reducing obstacle detection time for both ground robots and drones (Barry, Florence et al. 2018, Lamini, Fathi et al. 2024). However, efforts to accelerate detection often come at the expense of accuracy. In the pursuit of faster processing times, many researchers have been compelled to disregard a substantial portion of image pixels, resulting in a notable reduction in detection precision.

This trade-off between speed and accuracy highlights the urgent need for a more balanced approach. To fully realize the potential of stereo-based obstacle detection, it is crucial to develop methods that not only expedite the detection process but also maintain high levels of accuracy. Achieving this balance is essential to ensure that mobile robots can navigate complex environments safely and effectively, without sacrificing critical detection quality for the sake of speed.

This research presents an innovative method for obstacle detection that effectively balances speed and accuracy. We propose a novel hybrid approach that leverages the strengths of both stereo and monocular techniques. By employing monocular methods for obstacle identification and stereo techniques for depth estimation, our approach optimizes efficiency while preserving high levels of precision. In contrast to traditional methods that compute depth information for individual pixels, the proposed approach operates at the object level, calculating the distance for detected objects rather than processing each pixel independently.

Recent advancements in deep learning have made detecting objects from a single monocular image both fast and reliable (Qiu, Zhao et al. 2020, He and Liu 2021, John and Mita 2021, Lee, Ho et al. 2021). Since obstacles are physical entities in the real world, treating them as such during detection offers a more intuitive and effective approach. Our method enhances traditional techniques by focusing on image objects rather than individual pixels. This enables direct distance calculation to each object, eliminating the need for pixel-by-pixel processing. As a result, our approach significantly improves both the speed and accuracy of stereo-based obstacle detection, overcoming the limitations of prior methods. This represents a notable advancement in optimizing real-time obstacle detection for mobile robots, thereby enhancing their safety and efficiency in dynamic environments.

This article is structured to provide a clear and comprehensive understanding of our approach. Section 2 outlines the step-by-step process of the proposed algorithm for obstacle detection, emphasizing its key innovations. In Section 3, we present and analyze the experimental results, offering valuable insights into the method's effectiveness. Finally, Section 4 concludes with a summary of our findings and recommendations for future research, along with potential avenues for further improvement.

2. Methodology Framework

This section introduces the essential steps of the proposed obstacle detection technique, which has been meticulously designed to maximize both efficiency and precision. As illustrated in figure 1, each step plays a crucial role in optimizing performance and addressing key challenges in real-time obstacle detection. The following sections provide an in-depth analysis, demonstrating how this method not only enhances detection accuracy but also significantly improves processing speed, making it a robust and reliable solution for dynamic environments.

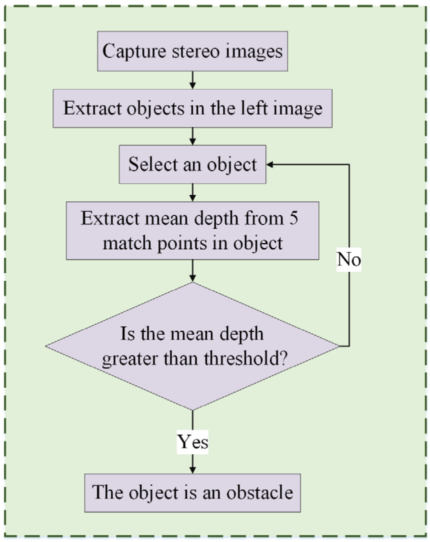

Figure 1. The overall structure of the proposed technique.

2.1. Data acquisition

In this study, we consider a stereo vision system consisting of two synchronized cameras mounted on the robot at a fixed baseline distance. This configuration captures two slightly different perspectives of the environment, facilitating depth estimation through disparity calculations. The captured image pair, referred to as the left and right images, provides complementary viewpoints that enable precise depth computation, thereby enhancing the reliability and robustness of the detection process. By utilizing both cameras, the proposed approach fully exploits the advantages of stereo vision, leading to improved obstacle identification and overall system performance. Figure 2 illustrates a stereo camera system mounted on a mobile robot, demonstrating its strategic placement for optimal depth perception and obstacle detection.

2.2. Extracting objects and depth calculating in object points

To achieve effective obstacle detection, we begin by selecting either the left or the right image, from which objects are extracted using advanced deep learning techniques. Specifically, we employ panoptic segmentation (Kirillov, He et al. 2019), a state-of-the-art approach that not only identifies countable objects but also captures amorphous, uncountable entities (Li, Qi et al. 2020). This comprehensive capability surpasses traditional segmentation methods, providing a more holistic understanding of the scene (Petrovai and Nedevschi 2022)

Figure 2. A pair of stereo cameras mounted on mobile robot (Yaqoob and Bajwa 2024).

Panoptic segmentation provides a comprehensive understanding of a scene by simultaneously addressing object recognition and background classification, enabling a more detailed and accurate interpretation of visual data. This integrated approach is particularly critical for high-stakes real-world applications, such as autonomous driving, robotics, and medical imaging, where precision and real-time processing are essential (Kirillov, He et al. 2019). By unifying instance and semantic segmentation within a single framework, panoptic segmentation streamlines the processing pipeline, reducing reliance on multiple models and thereby enhancing both efficiency and scalability (Xiong, Liao et al. 2019). As a result, this technique plays a pivotal role in advancing intelligent systems and optimizing operational performance across various industries.

Several advanced deep learning networks have been developed for panoptic segmentation, including UPSNet, FPSNet, EPSNet, VPSNet, DenseBox , Panoptic-deepLab, and LPSNet (Xiong, Liao et al. 2019, de Geus, Meletis et al. 2020, Kim, Woo et al. 2020). These networks have been trained on various datasets, such as Cityscapes (Cordts, Omran et al. 2016), which contains traffic-related images with 8 thing-type and 11 stuff-type categories; COCO (Lin, Maire et al. 2014), with 80 thing-type and 91 stuff-type categories; and Mapillary Vistas (Neuhold, Ollmann et al. 2017), which includes 37 thing-type and 28 stuff-type categories.

To maximize the precision and real-time performance of the proposed obstacle detection method, it is crucial to select a network and dataset that are both highly effective and efficient in detecting a diverse range of environmental objects. The COCO dataset, recognized as one of the most challenging and widely utilized benchmarks for evaluating panoptic segmentation, is an optimal choice for this application. Furthermore, recent studies have identified networks such as Panoptic-DeepLab, FPSNet, LPSNet, and PanoNet as leading solutions for fast panoptic segmentation (Kirillov, He et al. 2019, Hong, Guo et al. 2021). Among these, Panoptic-DeepLab, when trained on the COCO dataset, has exhibited superior precision, making it a highly suitable candidate for this task.

Thus, by selecting COCO and leveraging Panoptic-deepLab, we ensure that our obstacle detection approach is not only fast but also highly accurate, capable of detecting a wide array of objects in complex environments. This combination guarantees the most effective solution for real-time, high-precision obstacle detection in mobile robots.

By leveraging panoptic segmentation, the proposed method achieves exceptional accuracy in detecting all relevant objects within the environment, ensuring a robust and comprehensive obstacle detection system. This advanced approach not only enhances real-time performance but also positions itself as one of the most sophisticated and reliable solutions available, thereby establishing a new standard for obstacle detection technology.

After extracting objects, we select an object. Once the object is selected, we extract at least five matching points from both the left and right images. These match points are strategically arranged: one point at the center and four surrounding points, ensuring a well-distributed and accurate representation of the object's structure. To determine the key points, the convex hull of the extracted object is computed. The centroid of the object is estimated by calculating the average of the x and y coordinates of all points within the convex hull. For the selection of the surrounding points, the following criteria are applied: the point with the minimum x-coordinate is designated as the first point, the point with the maximum x-coordinate as the second point, the point with the minimum y-coordinate as the third point, and the point with the maximum y-coordinate as the fourth point. This method ensures a robust and consistent identification of key points based on the geometric properties of the feature.

Assuming the images are rectified, correspondence search is performed using the Block Matching (BM) (Cuevas, Zaldívar et al. 2013) method. BM is a computationally efficient stereo matching technique that identifies corresponding points by comparing fixed-size image blocks within the rectified stereo pair. Due to its simplicity and low computational cost, BM is well-suited for real-time applications.

Using these matched points, the precise distance (depth) of the object from the camera or mobile robot can be computed using the depth estimation formula presented in Equation (1). This approach ensures reliable and accurate depth estimation, enhancing the robot's capability to perceive its environment effectively. By obtaining precise depth information, the system improves navigation and obstacle avoidance, contributing to more robust and autonomous operation.

$$Z = (f \times B) / d, \quad (1)$$

where f = focal length
B = baseline (distance between two cameras)
d = disparity

2.3. Obstacle detection

The average depth of the five matched points is computed, and if this value exceeds a predefined threshold, the object is classified as an obstacle. The detected obstacle is then incorporated into the final binary obstacle image. This process is systematically applied to every detected object, ensuring that no obstacle is overlooked. As a result, a highly accurate and comprehensive binary obstacle map is generated, providing a precise representation of the environment. This enables the robot to navigate with enhanced confidence and reliability, facilitating robust and efficient obstacle avoidance.

3. Experimental results and discussion

In this section, we provide a comprehensive analysis of the performance of the proposed algorithm. The implementation was carried out using the Python platform, chosen for its flexibility and efficiency in algorithm development. For the initial evaluation, we utilized stereo images from the well-established Middlebury Stereo Evaluation database (Scharstein, Hirschmüller et al. 2014) (Figure 3). The image average resolution was 1260 × 2560 pixels. These images were acquired from a frontal perspective within a controlled environment containing various experimental objects. The dataset offers a diverse and extensive collection of images, making it well-suited for assessing the robustness and accuracy of the proposed algorithm under realistic conditions. By employing this standardized benchmark, we ensure that our findings are both reliable and directly comparable to existing methodologies.

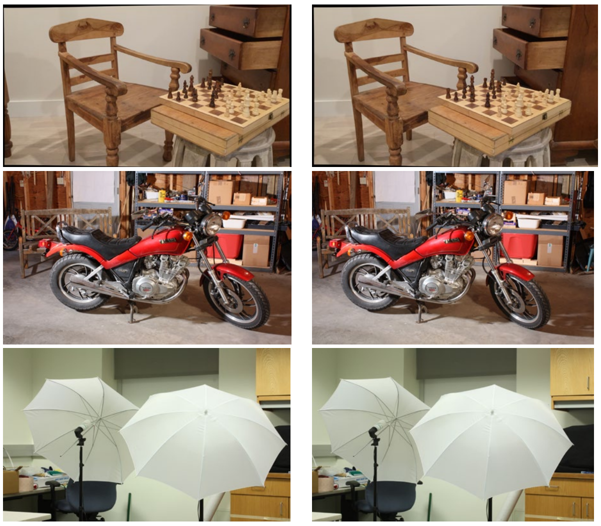

Figure 3. Sample images from the Middlebury Stereo Evaluation database.

Next, we used the Panoptic-deepLab network, trained on the challenging COCO dataset, to extract objects from the left image. The extracted objects presented in Figure 4.

Subsequently, the matching process is carried out using the Block Matching (BM). Notably, the Middlebury Stereo Evaluation database provides pre-rectified stereo image pairs, ensuring that corresponding points in the left and right images are precisely aligned along the same horizontal scanline. This rectification process enhances the accuracy and efficiency of stereo matching and depth estimation by eliminating the need for additional geometric corrections, thereby facilitating more precise and reliable results (Scharstein, Hirschmüller et al. 2014).

Subsequently, objects within a specified threshold distance were classified as obstacles. The final results, presented in figure 5, clearly differentiate obstacles from non-obstacles, with obstacles highlighted in white and non-obstacles in black. To accurately define the threshold distance for obstacle detection, we conducted a thorough analysis of key operational parameters. Taking into account the robot's speed of 10 meters per second and a minimum response time of 250 milliseconds, we determined that the robot requires at least 2.5 meters to detect and effectively respond to obstacles.

Figure 4. Objects extracted using the Panoptic-DeepLab network.

Thus, any object within this 2.5-meter range is classified as an obstacle, ensuring a prompt and reliable reaction to prevent collisions. This threshold calculation optimizes both safety and navigation efficiency, allowing the robot to operate with enhanced precision in dynamic environments.

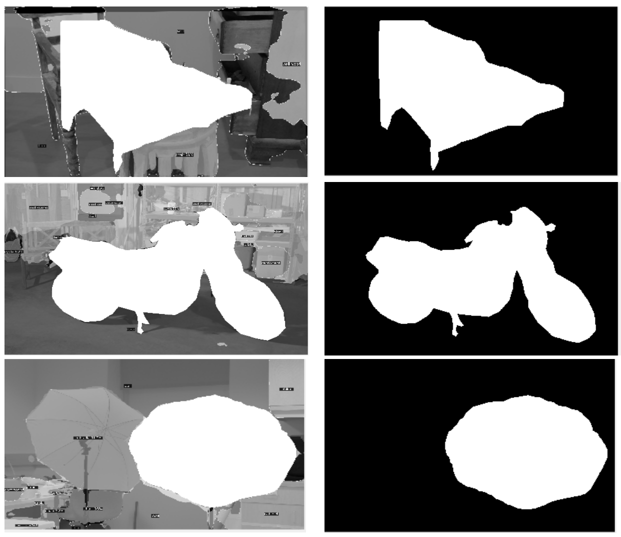

Figure 5. Obstacles identified using the proposed method.

To rigorously assess the proposed methodology, depth images derived from prior research (Lamini, Fathi et al. 2024) were employed, thereby establishing a robust foundation for our analysis. The evaluation was conducted using three critical parameters: recall, precision, and processing time, which enabled a comprehensive comparative analysis between our innovative approach and existing methodologies. The results of this evaluation are presented in Table 1. As illustrated in Table 1, the object detection not only significantly enhances the accuracy of obstacle detection but also expedites the detection process, thereby underscoring the effectiveness and superiority of our proposed method. This advancement holds substantial potential to enhance robotic navigation and safety within complex environments.

Parameters	The method proposed in this study	Existing method
Recall (%)	91	81
Precision (%)	85	56
Time (Ms)	180	300

Table 1. Comparison of the Results of the Proposed Algorithm and Existing Methods.

4. Conclusions

Stereo-based methods, while powerful, are often limited by their significant computational complexity. To reduce processing time, the search space typically needs to be constrained, either by lowering image resolution or excluding unnecessary segments, such as road areas in ground robots. However, these compromises can impact the effectiveness of obstacle detection. To overcome this challenge, we introduce a hybrid approach that seamlessly integrates the strengths of both stereo and monocular methods. Our proposed algorithm begins by extracting objects from the left image and identifying matching points to calculate their distances. If an object's distance exceeds a predefined threshold, it is classified as an obstacle. To evaluate the effectiveness of our method, we conducted extensive testing using front-facing images.

This innovative solution empowers obstacle detection by combining the speed and efficiency of monocular techniques with the precise depth estimation of stereo vision, achieving an optimal balance between accuracy and performance. By seamlessly integrating the strengths of both approaches and harnessing advanced deep learning techniques, we dramatically improve obstacle detection, making it more precise, adaptable, and scalable across diverse applications. The results are compelling—our method achieves an impressive accuracy rate of 91%, proving its reliability and effectiveness in real-world scenarios. This breakthrough paves the way for more intelligent, efficient, and robust obstacle detection systems.

Furthermore, stereo methods often face the challenge of precise camera calibration, a critical factor that can undermine the reliability of obstacle detection. Ensuring accurate calibration is essential for any stereo-based system to function properly. These complexities must be carefully addressed when developing a robust, real-time stereo-based detection algorithm.

References

Aguilar, W. G., Casaliglla, V. P. & Pólit, J. L., 2017. Obstacle avoidance based-visual navigation for micro aerial vehicles. *Electronics*, 6(1), p. 10.

Aharchi, M. & Kbir, M., 2022. Localization and Navigation System for Blind Persons Using Stereo Vision and a GIS. *WITS 2020*, Springer, 365-376.

Al-Kaff, A., García, F., Martín, D., De La Escalera, A. & Armingol, J. M., 2017. Obstacle detection and avoidance system based on monocular camera and size expansion algorithm for UAVs. *Sensors*, 17(5), 1061.

Badrloo, S. & Varshosaz, M., 2022. Detection of Obstacle Regions Around an MAV using an Expansion-based Technique. *Journal of Geomatics Science and Technology*, 11(3), 63-81.

Badrloo, S., Varshosaz, M., Pirasteh, S. & Li, J., 2022. Image-Based Obstacle Detection Methods for the Safe Navigation of Unmanned Vehicles: A Review. *Remote Sensing*, 14(15), 3824.

Barry, A. J., Florence, P. R. & Tedrake, R., 2018. High-speed autonomous obstacle avoidance with pushbroom stereo. *Journal of Field Robotics*, 35(1), 52-68.

Bharati, S. P., Wu, Y., Sui, Y., Padgett, C. & Wang, G., 2018. Real-time obstacle detection and tracking for sense-and-avoid mechanism in UAVs. *IEEE Transactions on Intelligent Vehicles*, 3(2), 185-197.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. & Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27-30 June.

Cuevas, E., Zaldívar, D., Pérez-Cisneros, M. & Oliva, D., 2013. Block-matching algorithm based on differential evolution for motion estimation. *Engineering Applications of Artificial Intelligence*, 26(1), 488-498.

de Geus, D., Meletis, P. & Dubbelman, G., 2020. Fast panoptic segmentation network. *IEEE Robotics and Automation Letters*, 5(2), 1742-1749.

Goodin, C., Carrillo, J., Monroe, J. G., Carruth, D. W. & Hudson, C. R., 2021. An Analytic Model for Negative Obstacle Detection with Lidar and Numerical Validation Using Physics-Based Simulation. *Sensors*, 21(9), 3211.

Grinberg, M. & Ruf, B., 2021. UAV Use Case: Real-Time Obstacle Avoidance System for Unmanned Aerial Vehicles Based on Stereo Vision. *Towards Ubiquitous Low-power Image Processing Platforms*, Springer, 139-149.

He, Y. & Liu, Z., 2021. A Feature Fusion Method to Improve the Driving Obstacle Detection under Foggy Weather. *IEEE Transactions on Transportation Electrification*.

Hong, W., Guo, Q., Zhang, W., Chen, J. & Chu, W., 2021. LPSNet: A lightweight solution for fast panoptic segmentation. *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 20-25 June.

John, V. & Mita, S., 2021. Deep Feature-Level Sensor Fusion Using Skip Connections for Real-Time Object Detection in Autonomous Driving. *Electronics*, 10(4), 424.

Kim, D., Woo, S., Lee, J.-Y. & Kweon, I. S., 2020. Video panoptic segmentation. *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 13-19 June.

Kirillov, A., He, K., Girshick, R., Rother, C. & Dollár, P., 2019. Panoptic segmentation. *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 15-20 June.

Lamini, C., Fathi, Y., Ba-ichou, A., Benhlima, S. & Bekri, A., 2024. New obstacle detection method for a mobile robot based on improved disparity map processing. *Journal of King Saud University - Computer and Information Sciences*, 102231.

Lee, H., Ho, H. & Zhou, Y., 2021. Deep Learning-based Monocular Obstacle Avoidance for Unmanned Aerial Vehicle Navigation in Tree Plantations. *Journal of Intelligent & Robotic Systems*, 101(1), 1-18.

Li, Q., Qi, X. & Torr, P. H., 2020. Unifying training and inference for panoptic segmentation. *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 13-19 June.

Lin, T.-Y. et al., 2014. Microsoft COCO: Common objects in context. *Proc. European Conf. on Computer Vision (ECCV)*, Zurich, Switzerland, 6-12 September, Springer.

Liu, J., Li, H., Luo, J., Xie, S. & Sun, Y., 2021. Efficient obstacle detection based on prior estimation network and spatially constrained mixture model for unmanned surface vehicles. *Journal of Field Robotics*, 38(2), 212-228.

Neuhold, G., Ollmann, T., Rota Bulo, S. & Kontschieder, P., 2017. The mapillary vistas dataset for semantic understanding of street scenes. *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, Venice, Italy, 22-29 Oct.

Petrovai, A. & Nedevschi, S., 2022. Fast Panoptic Segmentation with Soft Attention Embeddings. *Sensors*, 22(3), 783.

Qiu, Z. et al., 2020. Vision-based moving obstacle detection and tracking in paddy field using improved YOLOv3 and deep SORT. *Sensors*, 20(15), 4082.

Scharstein, D. et al., 2014. High-resolution stereo datasets with subpixel-accurate ground truth. *Pattern Recognition: 36th German Conference, GCPR 2014, Münster, Germany, September 2-5, 2014, Proceedings*, Springer.

Shi, T.-W. et al., 2023. Brain computer interface system based on monocular vision and motor imagery for UAV indoor space target searching. *Biomedical Signal Processing and Control*, 79, 104114.

Sun, B. et al., 2021. Obstacle Detection of Intelligent Vehicle Based on Fusion of Lidar and Machine Vision. *Engineering Letters*, 29(2).

Tijmons, S. et al., 2017. Obstacle avoidance strategy using onboard stereo vision on a flapping wing MAV. *IEEE Transactions on Robotics*, 33(4), 858-874.

Vargas, J. et al., 2021. An overview of autonomous vehicles sensors and their vulnerability to weather conditions. *Sensors*, 21(16), 5397.

Wu, J. et al., 2018. Study of multiple moving targets' detection in fisheye video based on the moving blob model. *Multimedia Tools and Applications*, 1-20.

Xiong, Y. et al., 2019. UPSNet: A unified panoptic segmentation network. *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 15-20 June.

Yaqoob, I. & Bajwa, I. S., 2024. Performance evaluation of mobile StereoNet for real-time navigation in autonomous mobile robots. *Multimedia Tools and Applications*, 83(12), 35043-35072.

Zhang, J. et al., 2023. Automated guided vehicles and autonomous mobile robots for recognition and tracking in civil engineering. *Automation in Construction*, 146, 104699.