Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 46 No. 2 (2025)

Vision-Based Framework for Autonomous Driving Using Stereo Vision

Mohammed Hamdan J*1, Dr. Vishwanath Koti2
Department of Mechanical Engineering, Ramaiah Institute of Technology, Bengaluru-54
Corresponding Author: Mohammed Hamdan

Abstract: - *Autonomous navigation in organized farming environments like spring-up fields of force and vineyards requires precise depth perception, obstacle avoidance, and waypoint-based life support. This work presents a vision-based architecture for autonomous operation of a ZED disco biscuit stereo camera, which performs literal prison term astuteness perception and object espial to facilitate safe and efficient piloting. In contrast to 3D LiDAR, which only records information on an exclusive plane without spatial awareness, or the costly 2D LiDAR, stereo visual modality has a left behind more efficient option for perceiving depth. Jibe (coinciding localization and chromosome mapping) makes an effective system pay the cost of formulating a real-time environmental map allowing accurate localization principle, while the Nav2 model simplifies sovereign waypoint adoption towards contacting navigation target area. As the car drives through craw words or vineyards, it continuously keeps guard over its charted flight, gripping directly if a man or barrier comes into its way of life and resuming apparent motion once impedimenta are drawn in; this allows for beneficial base interaction in dynamic environments. This approach performs superior adaptability and efficiency over LiDAR-based alternatives by using two-channel vision for depth selective information piloting, earning it an affordable and virtual solution for precision Department of Agriculture, independent husbandry, and agricultural robotic coating.*

Keywords*:Stereo vision, LIDAR sensors, Autonomous navigation, Object detection, SLAM.*

1. Introduction

The quick development of autonomous technology in agriculture has yielded novel solutions that seek to maximize efficiency, minimize human labor, and improve agricultural practice. Of the big challenges of automating farming, one stands out: the creation of affordable, scalable, and trustworthy self-navigating systems, especially for organized farming environments like row crops and vineyards. Conventional automation methods have mostly depended on expensive sensor technologies such as LiDAR and SONAR for environmental sensing and navigation. These technologies, however, have significant drawbacks that limit their application in large-scale agricultural implementations. LiDAR technology, though effective within robotic perception for many areas, has some profound disadvantages when implemented in agricultural areas. 2D LiDAR sensors measure on one plane, and because of this, they have drastically reduced capacity for perceiving depth and traveling across complex spaces of different heights. While 3D LiDAR sensors overcome this with superior spatial perception, their very expensive nature renders them unsuitable for most applications of agriculture, particularly in economically savvy farming practices. Likewise, SONAR-based solutions fail in rural environments because they are ineffective in highly dynamic environments with heavy vegetation, dust, and irregular terrain. To meet these challenges, we suggest a vision-based autonomous piloting system, tailor-made for structured agrarian landscapes, using stereo vision technology as a cost-effective and flexible option. Stereo cameras like the ZED X offer high-fidelity depth perception, strong object detection capabilities, and precise visual odometry, facilitating efficient and intelligent autonomous navigation in structured agricultural fields. Through the incorporation of Simultaneous Localization and Mapping (SLAM) and the Nav2 framework, our system guarantees end-to-end environmental awareness, accurate localization, and transparent waypoint-based navigation. As the self-driving car travels over organized fields like vineyards or row crops, it keeps to a predefined path while adaptively modifying its path based on real-time conditions. One important aspect of our system is adaptive obstacle detection and avoidance. When there is an obstruction, a person or a sudden agricultural object, in the way of the vehicle, the system stops automatically and moves again when the blockage is removed. While LiDAR tends to fail in dusty conditions or dense vegetation areas, stereo vision has real-time adaptive depth perception, which makes it very appropriate for rugged agricultural landscapes.

The independence of this vision-guided system poses a disruptive opportunity for contemporary precision agriculture, solving major challenges. With the help of real-time SLAM algorithms and vision-based sensing, the system can correctly map its environment, localize itself in the space, and move effectively through active and unstructured environments. The Nav2 platform supports goal-oriented waypoint navigation that allows the vehicle to navigate independently along specified paths, dynamically modify its path for obstacles, and arrive at defined destinations without intervention. In addition, the implementation of AI-powered object detection and classification greatly boosts situational awareness, allowing the system to discriminate between crops, obstacles, and unwanted objects and thus optimize route planning and navigation. This study seeks to redefine farm automation through the provision of an affordable, intelligent, and scalable solution that overtakes conventional sensor-based guiding mechanisms. Through the combination of stereo vision for depth perception, SLAM-based mathematical modeling for space representation, and AI-based decision-making algorithms, we create a very flexible and cost-effective navigation system. This method ensures both effectiveness and safety, making the move towards full autonomous farm operations smoother. With its cutting-edge autonomy, our system enables precision agriculture, optimizes productivity, and reduces dependence on labor-intensive farming methods, a major leap toward a more sustainable, smart, and automated future for contemporary farming.

Figure 1:ZEDX camera

2. Ease of Use

Stereo vision-based navigation takes precedence over LiDAR in agricultural robotics because of its affordability, better depth perception, and integrability with AI. Stereo cameras offer stable real-time depth mapping and obstacle detection compared to LiDAR, which fares poorly in dusty, foggy, and dense vegetation environments. Vision-based solutions facilitate AI-enabled crop health monitoring, visual odometry, and 3D obstacle detection at 120Hz with high-resolution 1080p imaging to improve situational awareness. The integration of AI maximizes adaptability with accurate path planning and autonomous navigation in organized farmlands. This method provides an effective, scalable, and robust alternative to conventional sensor-based agricultural automation techniques.

3. Literature Survey

Stereo vision is an important element in autonomous driving systems, providing precise depth perception and 3D environmental awareness through the utilization of two cameras. It supports crucial functions like obstacle detection, lane tracking, and path planning, making it a cost-effective and dependable substitute for other sensing technologies. This literature review discusses the latest studies and advances in vision-based systems with stereo vision, with an emphasis on their methodologies, challenges, and applications in autonomous driving. The following are the literature reviews pertinent to the subject matter. SS. Han et al., in their work [1] "A Guidance

Directrix Approach to Vision-Based Vehicle Guidance Systems", present the issue of accurate control of autonomous farm vehicles by using guidance information from crop rows. For image segmentation, an efficient approach was presented by utilizing K-means clustering and row detection with a moment algorithm followed by guidance line selection using a cost function. It was tried in corn and soybean fields with the RMS offset error being 1.0 cm in soybean and 2.4 cm in corn. It can be improved in multiple manners, i.e., inclusion of sensors to monitor complicated field conditions, boosting accuracy in non-ideal illumination conditions, and reducing computational loads for enabling real-time processing. Pengcheng Wei et al. [2] address the issues of high investment costs and low flexibility in current robot navigation systems. For edge detection in trajectory planning, a Canny-based ant colony algorithm is introduced, and for detection of obstacles, an SSD neural network is employed. The model was tested in simulation and had navigation accuracy of 89.62% and obstacle detection accuracy of 92.90%. This model is of value for applications, and future work may seek to enhance the speed of navigation and reduce energy consumption in challenging environments. Mohd Zishan et al. [3] refer to the limitations of traditional navigation systems in Automated Guided Vehicles, specifically in accuracy and flexibility of obstacle avoidance. They propose a vision-based AGV with a camera and ultrasonic sensors for obstacle detection and collision-free path guidance. Navigation is carried out using line detection algorithms. Future attention could be in terms of improving accuracy in obstacle recognition and system adaptability in changing environments. Josiah Radcliffe et al. [4], in their paper "Machine Vision for Orchard Navigation", discussed the problems of navigating orchards by farm robots utilizing machine vision to cope with conditions of variability. The system steers an unmanned ground vehicle between rows of an orchard using a detection technique of contrast between sky-tree canopy. A modified camera snaps photos which are pre-processed to guide navigation. Future enhancements could entail enhancing the robustness of the system in multiple environments and its use in multiple crops and soil types. Andrew English et al. [5] in their paper "Vision-Based Guidance for Robot Navigation in Agriculture" present a monocular vision-based texture-tracking technique for the navigation guidance of autonomous agricultural vehicles in fields. The system is capable of operating efficiently under visually challenging conditions and does not require crop-specific information. Future studies could attempt to enhance the robustness and adaptability of the method across various agricultural conditions and incorporate additional sensors to increase precision. Huei-Yung Lin et al. [6] introduce a vision-based autonomous navigation and collision avoidance system for quadrotors in low-altitude flight zones without employing gigantic sensors like LiDAR. The system is made up of a combination of real-time depth estimation and obstacle avoidance with optical flow techniques and pre-planned path planning by a Rapidly-exploring Random Tree (RRT) algorithm. Onboard cameras and IMUs are used for navigation and collision avoidance. Future improvements can include the refinement of obstacle detection accuracy and support for more complex path management across different conditions. M. K. Kaiser et al. [7], "Vision-Based Estimation for Guidance, Navigation, and Control of an Aerial Vehicle, 2010", propose a vision-based navigation of an aircraft based on features inside and outside the camera view of a monocular camera. The system demonstrates robust performance in GPS-denied environments with simulations and tests. Vision-based pose estimation error propagation can be an area of future research, along with sensor fusion with IMUs and GPS to offer bounds on errors in estimates under nominal conditions. Lounis Chermak et al. [8] introduce real-time vision-based navigation for space missions without GPS signals. The stereo vision system is combined with an Inertial Measurement Unit (IMU) to estimate ego-motion using a visual odometry algorithm. Feature tracking is assisted by IMU data on a small-form single-board computer. Future efforts can be focusing on processing efficiency and robustness improvements with state-of-the-art sensor fusion techniques for challenging environments. Aghi et al. [9] introduce a power-conscious, low-cost local motion planner for autonomous vineyard driving from solely an RGB-D camera and low-range components. The architecture utilizes stereo vision as depth maps for linear and angular velocities control of an unmanned ground vehicle (UGV). Inclement weather days on which depth maps may not function are addressed using a CNN-based steering mechanism. The planner is ROS-integrated, real-time-control capable, and model-pruned and quantized for optimization. Experiments in the field in northern Italy demonstrated its efficacy under different conditions. The strategy could be applied to orchards and other agriculture environments in the future. Miyamoto et al. [10] focus on the development of a visual navigation system for robots that mimic human navigation in city environments. In contrast to accurate metric maps, the proposed system relies on topological maps using landmarks. Image processing and semantic segmentation are paired to identify movement targets. In a university setting, a test track of 500 meters was conducted with a robot named "Emu" that used three webcams as sensors. With pedestrians, the robot successfully completed the track, proving the effectiveness of the conceived visual navigation strategy. Guangxu Liu et al. [11] present a detailed evaluation of vision-based manipulator pose measurement systems for autonomous hydraulic excavators. Having read 68 papers, they categorize vision applications under excavator detection, action identification, and posture estimation. The primary issues are speed, precision, dependability, and processing time. Marker-based solutions are easy but suffer from visibility issues, whereas deep learning eliminates the need for explicit modeling. Future research should focus on improving marker robustness, improving mapping models, building fault-tolerant algorithms, and expanding camera networks to offer broader views. Asadi et al. [12] create a mobile robotic platform for automation and data collection in construction sites. It is a vision-based system with the aim of streamlining operations and enhancing automation in difficult construction environments.With the application of advanced computer vision methods such as object detection and SLAM, the system enables accurate obstacle detection and movement in changing environments. Through the combination of GPUs and embedded boards, operations can be carried out independently, significantly reducing the need for human intervention and enhancing security. The effectiveness of the system is demonstrated by three case studies, which exhibit measurable productivity gains and reduced project timelines. Results indicate that the robotic platform is robust and flexible in that it is capable of adapting to various environmental conditions. Based on the findings of the research, this mobile robot system is an innovative advancement in construction technology with the potential to transform ways by improving workers' safety and efficacy. Hongkun Tian et al. [13] methodically explores the development and challenges of computer vision technology in agricultural robotics over the past three years. It highlights the significance of computer vision in enhancing small-field agriculture through the offering of low-cost, high-efficiency, and high-precision applications like crop growth observation, disease prevention, autonomous harvesting, quality inspection, automated farm management, and UAV monitoring. Notwithstanding such developments, challenges still exist, including the requirement for huge amounts of data, increased professional requirements, and ensuring stable operation in complex conditions. The authors assert that agricultural automation's future will involve integrating computer vision with smart technologies like deep learning, thereby enhancing the economic and functional performance of agricultural systems. This technology synergy aims to enhance efficiency and quality in farm production, providing valuable decision-support to farmers. D. Banerjee et al. [14] research outlines an integrated test automation system that was designed specifically to examine a motion-based image capture system utilizing a robotic arm. This new method bypasses the constraints of human control by automating the synchronization process using on-device test scripts, significantly enhancing testing accuracy and efficiency. The automation rules out the chance of human error and allows for consistent, reproducible test conditions, which are essential for evaluating the performance of the picture capture system. There is also a strong focus on enhancing the framework's scalability, in order to be able to effectively accept more diversified software product lines. The researchers also aim to explore new ways for enhancing the accuracy and efficiency of motion-based picture capturing systems. Xiaochun Luo et al. [15] research centers on workspace design, which is essential in coordinating individuals and groups to enhance performance and safety. It highlights the importance of acquiring and visualizing dynamic workplaces to effectively evaluate planning strategies. The authors report an initial attempt at applying state-of-the-art computer vision methods for the automatic identification and visualization of dynamic workplaces where workers on foot are present. The system acquires action data in two ways: action classes and action locations, through object detection, multiple object tracking, and action recognition. To process this data, a density-based spatial clustering method is employed to determine dynamic workspaces. The analysis of the integrated system's methods showed similar performance to original methods used in more general settings. Overall, this research is a significant contribution towards improving workspace planning using computer vision technologies. Central question gaps for vision-based farm automation systems involve the flexibility of the plan to a wide variety of both forecasted and unforecasted environments, e.g., light, weather, and crop types.

Therefore, these aspects need to be considered when detailing the strength of the system, especially in terms of preparedness in the field. The following gap is artificial intelligence and sensor integration. Technologies like deep learning, GPS, and extra cameras can also contribute to further enhancing decision-making and validity in real time. Navigation and obstacle detection need to be improved, especially towards the detection of moving objects and navigation over complex agricultural terrain, like orchards and non-planar fields. Scalability and hardware tuning are challenges awaiting solution in the scaling of such systems to bigger agricultural tasks since real-time response necessitates good utilization of the hardware. The final argument is that the scarcity of big-data datasets is preventing algorithms from attaining precision and robustness. Therefore, the aim would be to discover more effective algorithms that will support automation in farms, especially with the current missing features or shortage in environments.

4. METHODOLOGY

4.1 Stereo Camera Setup

This study explores object detection and depth estimation using stereo vision with the ZED X camera, focusing on two primary applications: human detection for collision avoidance and row crop lane following in agricultural environments.

4.1.1 ***Object Detection with ZED SDK***

The ZED SDK includes an integrated object detection module that can detect specific objects, e.g., people and cars, in the video stream. The module uses a very optimized AI model for detecting objects and, by tapping into the depth sensing of the camera, computes their 3D position in the world. This makes it possible to overlay virtual bounding boxes over actual objects, enabling applications such as collision avoidance in agriculture.

4.1.2 Custom Object Detection Integration

For more specific object detection applications, like detecting certain agricultural equipment or certain types of crops, the use of custom detectors is necessary. The ZED SDK facilitates the use of custom object detectors by consuming 2D bounding box detections from third-party models and calculating their 3D locations based on the depth information of the camera. This allows the system to recognize a broader variety of objects than the default classes the ZED SDK supports.

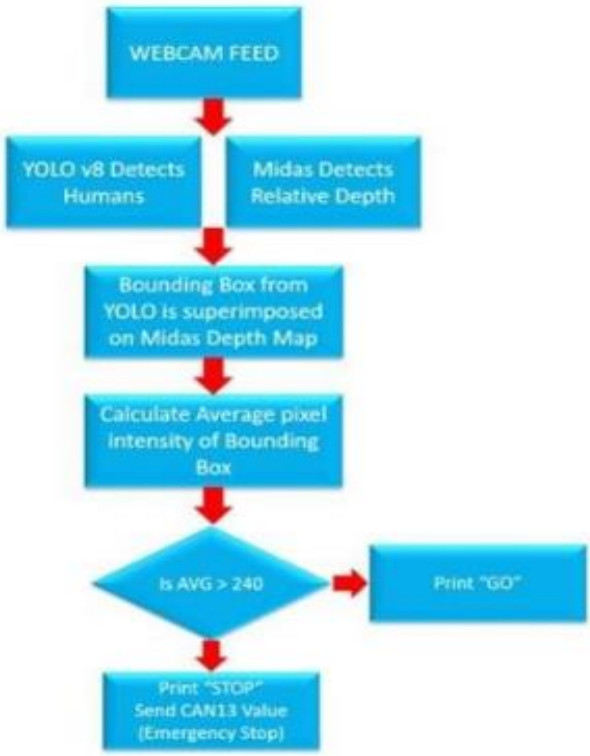

Figure 2 :Object Detection Flow Chart

4.1.3 Row crop lane following

In organized farm conditions, correct alignment within rows of crops is imperative. Depth information from the ZED X camera can be used to detect edges of crop rows so that the system can correctly follow lanes. Through ongoing inspection of the depth map, the system can spot drift from row center and align its path correspondingly to enable accurate passage across the fields.

4.1.4 Advantages of Stereo Vision over Monocular Systems

The ZED X stereo camera provides distinct benefits compared to monocular depth estimation models by offering absolute depth readings, increasing the reliability for real-world applications. This feature is especially useful in agricultural environments, where precise distance measurement is critical for functions such as obstacle avoidance and row following. The stereo vision systems are also typically less expensive than LiDAR-based systems, providing an economical alternative for mass deployment in agriculture. By taking advantage of the stereo vision capabilities of the ZED X camera, this approach is intended to improve the precision and dependability of object detection and navigation in farm settings and thereby enable safer and more efficient autonomous farming practices.

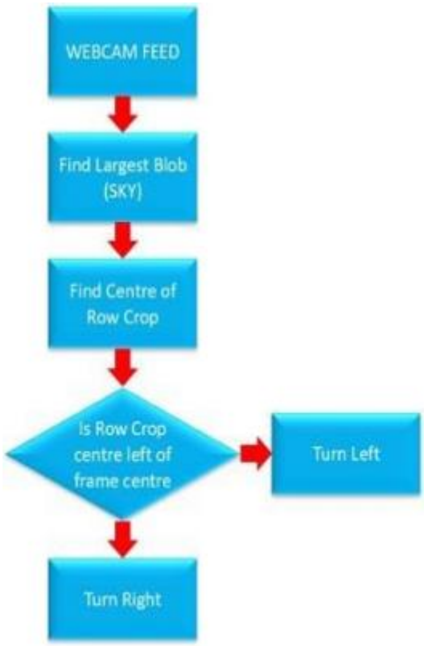

Figure 3: Row crop lane keeping architecture

4.2 Stereo Camera setup

In our experimental configuration, a ZED X camera is attached to a Jetson Orin Nano development kit using a GMSL2 capture card, with a 1-to-4 FAKRA cable for the camera and two CSI ribbon cables for connecting to the Jetson. The capture card is powered by a 2A, 12V power supply, while the Jetson Orin Nano is powered by an external 9V to 12V power source to provide high-bandwidth data transmission required for real-time autonomous navigation.

Stereo cameras such as the ZED X work by taking pictures from two slightly dissimilar viewpoints, replicating human binocular vision.

By computing the difference between such images, the system approximates object distances, which is very important for robotics and autonomous vehicle applications like object detection, obstacle avoidance, and terrain mapping.

The mathematical concept of stereo vision relies on triangulation, in which disparity computations between corresponding stereo image points create a depth map, allowing for a three-dimensional representation of the environment.

4.2.1 Triangulation in Stereo Vision

Triangulation is the essence of stereo vision and determines the 3D location of points from their 2D locations in the left and right images. For a point P in the world, its 2D locations in the left and right cameras are P1 and P2, respectively. The disparity d is the horizontal displacement between P1 and P2. The depth Z is calculated by the following equation. Z = f.B/d Where: • f is the focal length of the cameras. • B is the baseline, the distance between the two cameras. • d is the disparity.

4.3 Disparity Calculation

Disparity is the x-distance between matching points in the two images. In rectified stereo images, disparity d is given by the difference in x-coordinates between matching points in the left and right images. d = $X1 - X2$ Where: X1 and X2 are the x-coordinates of corresponding points in the left and right images, respectively. Closer objects exhibit a larger disparity, and further objects exhibit a smaller disparity.

4.4 Creating a Depth Map

After computing the difference d for every pixel of the stereo images, a depth map is produced. The depth map is an image in grayscale or color with color-coded pixel intensities, which correspond to the depth of the respective 3D points. The depth Z is computed with the previously mentioned formula. Z = f.B/d Where: • f is the focal length of the cameras. • B is the baseline, the distance between the two cameras. • d is the disparity. In the depth map, closer objects appear brighter, and further objects appear darker.

5.Results and Analysis

5.1 Object Detection

5.1.1 Using only YOLOv8

In Fig. 4, the limitations of employing YOLOv8 individually for object detection are obvious. In Fig. 4.a, the system prints "GO" if the BB takes up less than 20% of the screen space.

In Fig. 4.b, however, when the BB takes up more than 20%, the system misprints "STOP" due to posture and height variations, for example, when a person outstretches their arms standing far away. This discrepancy indicates the shortfalls of relying purely on BB area when making decision

Figure 4: Disadvantage of Object Detection using YOLO

5.1.2 Using YOLOv8 and MiDaS

In order to tackle this issue, MiDaS depth model was incorporated with YOLOv8, as displayed in Fig. 5.a and Fig. 5.b, showing better precision. By comparing Fig. 4.a with Fig. 5.a, and Fig. 4.b with Fig. 5.b, it is apparent that incorporating depth reduces the inaccuracies due to height and posture. For the same scenarios, the overall YOLOv8 and MiDaS system correctly prints "GO" in both situations, as depth estimation is a more reliable metric for decision-making. The system thereby efficiently alleviates the disadvantages of using YOLOv8 on its own, improving the reliability of the system.

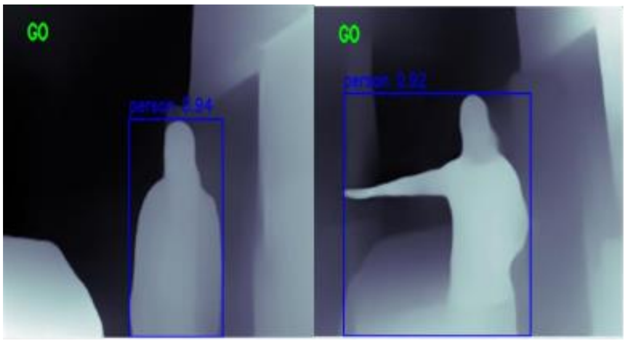

Figure 5: Overcoming the disadvantage using YOLO and MiDaS

Figure 6: Object Detection using YOLO and MiDaS

5.2 Row crop lane keeping

5.2.1 Using cv2 color segmentation

Fig.7 illustrates how the color segmentation process assists in maintaining the lane in crop rows. The method employs OpenCV to identify crop rows based on their colors and encloses bounding boxes around the left and right rows.

The process is straightforward and can be executed quickly, making it appropriate for applications in real-time. By measuring the lengths of the bounding boxes, the system determines the direction: "LEFT" or "RIGHT." This approach offers a simple and efficient method to track crop rows without the use of complicated sensors or sophisticated algorithms.

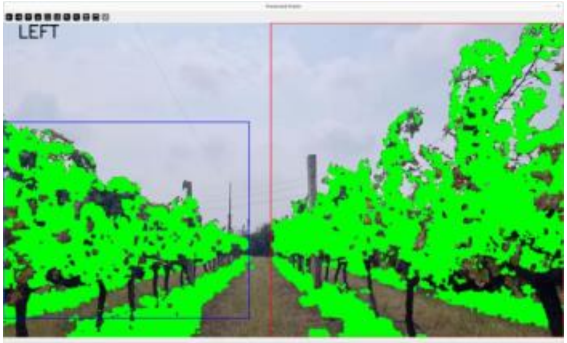

Figure 7:Row Crop Lane Keeping

5.2.2 Identifying perspective using largest blob

Fig.8 depicts the perspective-based row crop lane detection approach. Under this technique, the sky was determined as the largest area and was colored white, whereas everything else was colored black.

This segmentation properly accentuated the shortest black area, which symbolizes the crop row center. Since there is the natural effect of perspective, the center would take on a clear "V" shape with closer taller plants and distant shorter ones. The system properly identified row crop center displacements with respect to the frame center. It printed "RIGHT" when the center moved left and "LEFT" when the center moved right, showing a stable and easy means of lane guidance in farm environments.

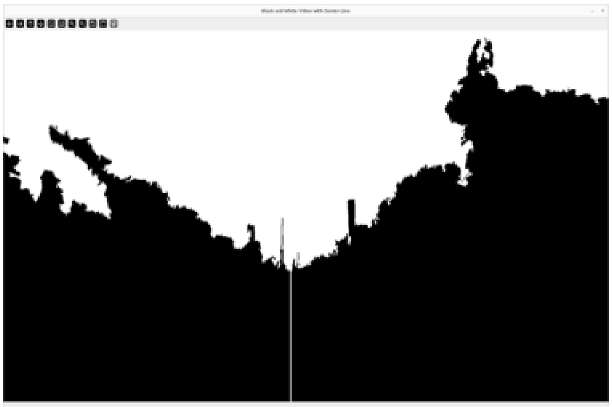

Figure 8: Row Crop lane keeping using Largest Blob to find perspective.

5.3 Stereo vision

5.3.1 Generation Depth map

Disparity and depth maps were calculated based on the principles of triangulation. The disparity was the horizontal pixel offset between corresponding left and right rectified image points.

This procedure yielded an absolute depth map where red pixels signified closer objects and blue pixels indicated distant objects. The depth map gave accurate spatial comprehension, as indicated by Fig.9.

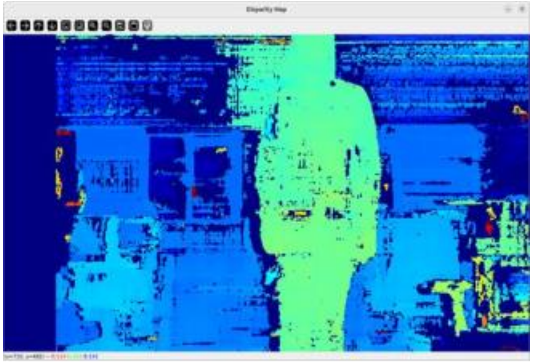

Figure 9:Depth Map with Stereo Camera.

3.2 Human detection and distance calculation

The YOLOv8 model was utilized for the detection and distance estimation to humans. The model detected the individuals in the scene and identified their bounding boxes. The distance to every detected person was calculated from the depth map by the stereo camera system, where the distance to the closest detected person was used for further decision. In case the detected person's distance is less than a specified threshold, the system can be commanded with a "STOP" command. Otherwise, it may be signaled "GO" like in Fig.10.

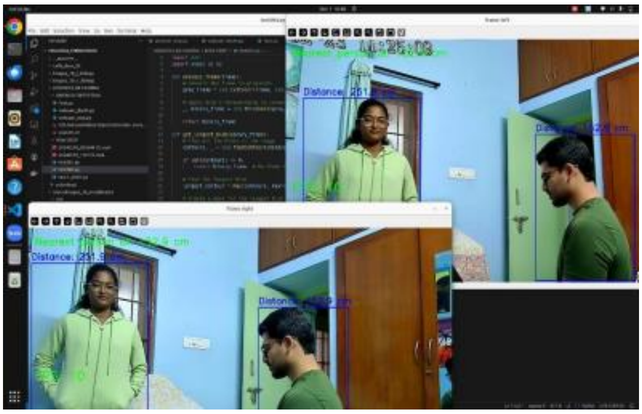

Figure 10:Human Detection and distance calculation using stereo camera

5. Conclusion and Future Work

The envisioned vision-based automation system overcomes the shortcomings of conventional sensor-based approaches to agricultural and warehouse automation successfully through an affordable, scalable, and responsive option.

The system utilizes the ZED X stereo camera to access depth estimation, visual odometry, and spatial mapping, thus delivering accurate object detection, row crop lane following, and distance measurement through pure vision-based methods. The combination of ZED X's stereo vision with sophisticated depth perception algorithms provides dynamic navigation in adverse conditions and less reliance on costly sensors such as LiDAR.

The reliability of the system under varied operating conditions emphasizes its applicability and sustainability for contemporary farming and logistics use. For more advanced performance, SLAM methods like ORB-SLAM3 can be combined with the ZED X stereo camera to provide more robust localization through tracking of keypoints and real-time map building.

Sensor fusion of IMU measurements can be used to enhance motion estimation and compensate for drift in visual odometry to provide stable operation in dynamic scenes. The point cloud generation ability of the system can be utilized for accurate 3D scene reconstruction, facilitating improved obstacle detection and path planning. With integration into ROS2's Nav2 stack, the system can be used to perform autonomous navigation, where robots can make real-time decisions, dynamically plan paths, and navigate row crops or warehouse spaces efficiently. Subsequent optimizations will address algorithmic performance, real-time execution on edge devices such as Jetson Orin Nano, and increasing datasets to enhance versatility across different terrain and lighting environments. These developments will lead to a more accurate, scalable, and resilient vision-based automation system that can work reliably in complex agricultural and warehouse settings.

References

[1] Han, S., Zhang, Q., Ni, B. and Reid, J.F., 2004. A guidance directrix approach to vision- based vehicle guidance systems. Computers and Electronics in Agriculture, 43(3), pp.179- 195.

[2] Wei, P., Yu, X., Di, Z., Dai, X., Wang, B. and Zeng, Y., 2022. Design of robot automatic navigation under computer intelligent algorithm and machine vision. Journal of Industrial Information Integration, 28, p.100366.

[3] Zishan, M. Modern Navigation System of Automatic Guided Vehicle (Vision-Based Navigation System).

[4] Radcliffe, J., Cox, J. and Bulanon, D.M., 2018. Machine vision for orchard navigation. Computers in Industry, 98, pp.165-171.

[5] English, A., Ross, P., Ball, D. and Corke, P., 2014, May. Vision based guidance for ro- bot navigation in agriculture. In 2014 IEEE International Conference on Robotics and Auto- mation (ICRA) (pp. 1693-1698). IEEE.

[6] Lin, H.Y. and Peng, X.Z., 2021. Autonomous quadrotor navigation with vision based ob- stacle avoidance and path planning. IEEE Access, 9, pp.102450-102459.

[7] Kaiser, M.K., Gans, N.R. and Dixon, W.E., 2010. Vision-based estimation for guidance, navigation, and control of an aerial vehicle. IEEE Transactions on Aerospace and Electronic Systems, 46(3), pp.1064-1077. [8] Chermak, L., Aouf, N., Richardson, M. and Visentin, G., 2019. Real-time smart and- standalone vision/IMU navigation sensor. Journal of Real-Time Image Processing, 16, pp.1189-1205.

[9] Aghi, D., Mazzia, V. and Chiaberge, M., 2020. Local motion planner for autonomous navigation in vineyards with an RGB-D camera-based algorithm and deep learning synergy. Machines, 8(2), p.27.

[10] R. Miyamoto et al., ”Vision-Based Road-Following Using Results of Semantic Semen- tation for Autonomous Navigation,” 2019, IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin), Berlin, Germany, 2019, pp. 174-179, doi:10.1109/ICCE Ber- lin47944.2019.8966198.

[11] Guangxu Liu, Qingfeng Wang, Tao Wang, Bingcheng Li, Xiangshuo Xi, Vision-based excavator pose estimation for automatic control. Automation in Construction, Volume 157, 2024, 105162, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2023.105162.

[12] Asadi, Khashayar, Hariharan Ramshankar, Harish Pullagurla, Aishwarya Bhandare, Su- raj Shanbhag, Pooja Mehta, Spondon Kundu, Kevin K. Han, Edgar J. Lobaton and Tianfu Wu. “Vision-based integrated mobile robotic system for real-time applications in construc- tion.” Automation in Construction (2018): n. pag. [13] Hongkun Tian, Tianhai Wang, Yadong Liu, Xi Qiao, Yanzhou Li.Computer vision technology in agricultural automation — A review. Information Processing in Agriculture, Volume 7, Issue 1, 2020.

[14] D. Banerjee and K. Yu, ”Integrated Test Automation for Evaluating a Motion-Based Image Capture System Using a Robotic Arm,” in IEEE Access, vol. 7, pp. 1888-1896, 2019, doi: 10.1109/ACCESS.2018.2886272.

[15] Xiaochun Luo, Heng Li, Hao Wang, Zezhou Wu, Fei Dai, Dongping Cao, Vision-based detection and visualization of dynamic workspaces. Automation in Construction, Volume 104, 2019, Pages 1-13, ISSN 09265805, https://doi.org/10.1016/j.autcon.2019.04.001.

[16] D. N. Vinod and T. Singh, "Autonomous Farming and Surveillance Agribot in Adja- cent Boundary," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bengaluru, India, 2018, pp. 1-7, doi: 10.1109/ICCCNT.2018.8494137.

[17] D. N. Vinod and T. Singh, "Autonomous Agricultural Farming Robot in Closed Field," 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2018, pp. 1-7, doi: 10.1109/RTE- ICT42901.2018.9012118.

[18] E. Soumya, T. Singh and S. Santhanalakshmi, "An Intelligent System for Face Mask Recognition using Open Computer Vision and Modified YOLO Algorithm," 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 2022, pp. 1-5, doi: 10.1109/INDICON56171.2022.10040177.

[19] E. Soumya, T. Singh and S. Santhanalakshmi, "An Intelligent System for Face Mask Recognition using Open Computer Vision and Modified YOLO Algorithm," 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 2022, pp. 1-5, doi: 10.1109/INDICON56171.2022.10040177.

[20] Singh, T., Nair, R.R., Babu, T. and Duraisamy, P., 2024. Enhancing Academic Integrity in Online Assessments: Introducing an Effective Online Exam Proctoring Model using YOLO. Procedia Computer Science, 235, pp.1399-1408.

[21] Sugunadevi, C., Rimjhim Padam Singh, B. Uma Maheswari, and Priyanka Kumar. "Di- aMOS Plant Leaves Disease Classification using Vision Transformer." In 2023 9th Interna- tional Conference on Signal Processing and Communication (ICSC), pp. 551-556. IEEE, 2023. bpage: https://indiankanoon.org