International Journal of Informatics and Computation (IJICOM)
Vol. 7, No. 2, 2025
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Predicting Real Distance for Wheeled Soccer Robot using YOLO Architecture

Sholahuddin Muhammad Irsyad[1], Agus Khumaidi[2], Ryan Yudha Aditya[3], Adi Rahmad Ramadhan[4*], Dhika Arya Pratama[5*], Muhammad Jardin Saputra[6*]

Abstract

This study presents a real-distance estimation system for wheeled soccer robots that integrates stereo vision cameras with a YOLO-based object detection algorithm to support accurate perception for game strategies. Experimental evaluations using 10 ball samples and 6 robot scenarios demonstrate that the proposed system achieves an average distance estimation accuracy of 96.7% with a low error margin of ±2.3% and a confidence level of 99%, indicating high reliability in object detection and metric distance measurement. The results confirm that stereo vision combined with deep learning provides precise spatial information suitable for dynamic soccer robot environments, enabling improved positioning and decision-making. While the system performs robustly under standard conditions, future work will address performance degradation caused by lighting variations, explore newer YOLO model architectures, and incorporate artificial intelligence–based adaptive strategies to further enhance autonomy and competitiveness in wheeled soccer robot applications.

Keywords:
Stereo Vision, YOLOv5, ROS, Robot Soccer, Autonomous Navigation

1. Introduction

Robot soccer competitions, particularly the Indonesian Robot Contest (KRI), demand precise perception, fast decision-making, and accurate motion execution to enable effective game strategies under dynamic and adversarial conditions. Wheeled soccer robots must continuously perceive the positions of teammates, opponents, and the ball while operating within strict competition rules and real-time constraints. Among these requirements, accurate distance estimation plays a crucial role in determining passing strength, shooting decisions, obstacle avoidance, and formation control. The official KRI guidelines emphasize autonomous perception and intelligent behavior as core evaluation criteria, yet many teams still struggle to achieve reliable real-world distance prediction in cluttered and fast-changing environments. This challenge highlights the need for robust vision-based distance estimation methods that can operate consistently during matches and directly support tactical decision-making [1].

Recent advances in object detection using deep learning significantly improve visual perception in mobile robotics. YOLO-based architectures, such as YOLOv4 and YOLOv5, demonstrate strong performance in detecting dynamic objects with low latency, making them suitable for robot soccer applications. Studies on wheeled mobile robots show that

Corresponding Author: Sholahuddin Muhammad Irsyad (muhammad.irsyad@ppns.ac.id)
1 Sholahuddin Muhammad Irsyad, Politeknik Perkalapan Negeri Surabaya (muhammad.irsyad@ppns.ac.id)
2 Agus Khumaidi, Politeknik Perkalapan Negeri Surabaya (aguskhumaidi@ppns.ac.id)
3 Ryan Yudha Aditya, Politeknik Perkalapan Negeri Surabaya (ryanyudhaadhitya@ppns.ac.id)
4 Adi Rahmad Ramadhan, Politeknik Perkalapan Negeri Surabaya (adirahmad@student.ppns.ac.id)
5 Dhika Arya Pratama, Politeknik Perkalapan Negeri Surabaya (dhikaarya12@student.ppns.ac.id)
6 Muhammad Jardin Saputra, Politeknik Perkalapan Negeri Surabaya (muhammadjardin@student.ppns.ac.id)

improved YOLO variants enhance detection accuracy and robustness under varying lighting and motion conditions. However, most object detection studies focus primarily on classification and localization in image space, without explicitly addressing real-world distance prediction. As a result, robots often detect objects accurately but still rely on heuristic or indirect methods to estimate distance, which limits the effectiveness of strategic actions that depend on spatial accuracy [2].

To overcome the limitations of monocular vision, researchers increasingly adopt stereo vision systems to estimate depth directly from visual disparity. Stereo cameras enable robots to infer real-world distances by exploiting geometric relationships between paired images. Prior work on wheeled soccer robots demonstrates that combining feature extraction methods, such as Harris corner detection, with YOLO-based object detection can improve distance estimation performance. These approaches show promising results in controlled environments, but they still face challenges related to calibration errors, computational load, and sensitivity to environmental noise. Consequently, achieving stable and accurate distance prediction during competitive matches remains an open research problem [3].

Depth-sensing technologies, including RGB-D and stereo-based systems, further illustrate the importance of accurate depth information in robotic perception. Studies using depth cameras, such as the Orbbec Astra and Kinect sensors, confirm that depth data significantly enhances spatial understanding for tasks like fall detection and motion analysis. While these sensors provide reliable depth measurements, their cost, power consumption, and susceptibility to outdoor lighting conditions limit their applicability in robot soccer competitions. Stereo vision offers a more flexible and competition-friendly alternative, but it requires careful algorithm design to match the accuracy of dedicated depth sensors [4], [7].

Beyond robotic perception, deep learning-based detection frameworks continue to evolve, with newer YOLO versions demonstrating improved speed and accuracy across diverse applications. Research on traffic flow estimation using YOLOv7 highlights the scalability and real-time capabilities of modern object detection networks. These strengths are directly relevant to robot soccer, where multiple objects move rapidly, and decisions must occur within milliseconds. Nevertheless, most studies emphasize detection metrics such as precision and recall, while depth accuracy and distance consistency receive less attention. This gap underscores the need to integrate detection performance with reliable spatial estimation for strategic robotic behavior [5].

The integration of the Robot Operating System (ROS) with vision-based perception further enhances modularity and real-time processing in soccer robots. Prior implementations show that ROS-based pipelines improve ball detection accuracy and system coordination when combined with YOLO models. However, these systems often treat distance estimation as a secondary process or rely on simplified geometric assumptions. As a result, robots may detect objects accurately but fail to compute precise distances needed for coordinated team strategies, such as synchronized passing or defensive positioning [6].

Strategic decision-making in wheeled soccer robots strongly depends on accurate spatial awareness. Research on decision tree-based strategies and gyrodometry-trigonometry methods demonstrates that precise position and distance information directly influence shooting angles, passing success, and tactical effectiveness. When distance estimation errors occur, strategic models produce suboptimal decisions, leading to missed passes or inefficient movements. These findings reinforce the importance of integrating accurate stereo-based distance prediction into higher-level strategy modules to ensure that perception directly supports intelligent decision-making [9], [12].

Foundational studies in stereo vision and multi-view geometry provide the theoretical basis for real-world distance estimation in robotic systems. Classical works on stereo correspondence, camera geometry, and visual odometry establish robust mathematical frameworks for depth estimation. More recent research extends these principles to real-time robotics applications, addressing challenges such as illumination changes, motion blur, and computational efficiency. Despite these advances, applying stereo vision reliably in robot soccer environments remains challenging due to rapid motion, occlusion, and limited processing resources. Therefore, a focused investigation into predicting real distance using stereo vision, tailored specifically for wheeled soccer robot game strategies, remains both relevant and necessary [16]–[21].

2. Related Works

Several studies investigate deep learning–based object detection as a foundation for perception in mobile and soccer robots. Hu et al. propose an improved YOLOv4 model for wheeled mobile robots and report detection precision exceeding 92% under structured indoor conditions, demonstrating strong robustness against motion blur and partial occlusion. This work highlights the advantage of single-stage detectors for real-time robotics; however, it focuses primarily on image-space localization and does not address real-world distance estimation. As a result, the detected object positions remain insufficient for strategic decision-making that depends on metric distance information [2].

Research specifically targeting robot soccer environments increasingly integrates object detection with stereo vision. Ramadhan et al. combine Harris corner detection with YOLOv5 on a stereo camera system for wheeled soccer robots and report an average distance estimation accuracy of approximately 90–93% within short to medium ranges. Their work demonstrates that stereo disparity significantly improves spatial perception compared to monocular approaches. Nevertheless, the method remains sensitive to camera calibration errors and lighting variations, which reduce accuracy during fast gameplay and dynamic match conditions [3].

Depth-based perception studies using RGB-D sensors provide important insights into distance reliability. Biswas et al. use Orbbec Astra 3D Pro depth images for fall detection and achieve detection accuracy above 95%, confirming the effectiveness of depth sensing for spatial reasoning. Although this approach delivers high accuracy, the reliance on dedicated depth hardware limits its applicability in robot soccer competitions, where stereo cameras are preferred due to cost, flexibility, and competition constraints. This limitation motivates further research into achieving comparable accuracy using stereo vision alone [4].

Advancements in YOLO architectures continue to influence robotic perception research. Saputri et al. apply YOLOv7 for traffic flow estimation and report detection accuracy above 94% with real-time performance. Their findings confirm that modern YOLO variants scale well to multi-object scenarios and high-speed environments. However, the study focuses on counting and classification tasks rather than spatial distance prediction, leaving a gap between detection performance and actionable spatial awareness required for soccer robot strategies [5].

The integration of Robot Operating System (ROS) with vision algorithms enhances modularity and real-time coordination in robotic systems. Wahyudi et al. demonstrate that combining ROS with YOLOv5 improves ball detection accuracy for wheeled soccer robots to approximately 91%, enabling more stable perception pipelines. Despite this improvement, the system still relies on indirect distance estimation methods, which limit its ability to support advanced tactics such as adaptive passing force or dynamic positioning based on precise opponent distances [6].

Strategic decision-making frameworks emphasize the importance of accurate distance information. Darmawan et al. design decision tree–based strategies for wheeled soccer robots and show that accurate spatial inputs significantly improve decision consistency, although they do not explicitly quantify distance accuracy in percentage terms. Their work highlights that even sophisticated strategy models underperform when perception inputs lack metric precision, reinforcing the need for reliable stereo-based distance estimation to support higher-level reasoning [9].

Stereo vision accuracy has also been evaluated independently of soccer contexts. Winarti et al. improve stereo distance measurement on humanoid soccer robots and report distance error reductions corresponding to accuracy levels of approximately 88–92% after calibration optimization. Similarly, Abdelsalam et al. analyze the ZED 2i stereo camera and show depth accuracy above 90% in controlled indoor environments. These studies confirm that stereo vision can achieve high metric accuracy, but they also reveal sensitivity to lighting, baseline configuration, and processing latency, which remain challenging in competitive robot soccer settings [11], [13].

Foundational and advanced works in stereo vision and visual odometry establish the theoretical and algorithmic basis for real-time distance estimation. Classical studies by Scharstein and Szeliski, as well as Hartley and Zisserman, define robust stereo correspondence and multi-view geometry frameworks, while more recent works, such as direct sparse odometry, demonstrate high-precision motion and depth estimation in real time. Although these approaches achieve high accuracy—often exceeding 90% in benchmark datasets—they typically assume stable environments and substantial computational resources. Adapting these methods to wheeled soccer robots requires simplification and optimization to balance accuracy, speed, and robustness during fast-paced matches [18]–[21].

3. Proposed Method

3.1 Orbbec Astra Pro Plus Stereo Camera

The Orbbec Astra Pro Plus wheeled soccer robot's stereo camera utilizes depth sensors to detect and measure the distance of surrounding objects, such as the ball and opposing robots [15]. The use of this depth data assists the robot in various functions, such as monitoring the ball, avoiding opposing robots, determining the direction of movement, and accurately picking up the ball. The robot's ability to adapt to changing game situations was tested under various field conditions [4]

Fig. 1. Orbbec Astra Pro Plus

To use the Orbbec Astra Pro Plus camera, you need support from the OpenNI interface and API. OpenNI helps develop natural interaction-based applications and acts as a bridge between data processing software and visual sensors such as the Orbbec Astra Pro Plus camera. Developers don't need to worry about compatibility between sensors and software with standard data formats, which is the main purpose of OpenNI. This method allows sensor manufacturers, such as Orbbec, to focus on hardware development, while application developers can use standard data outputs to create cross-platform applications. In addition, OpenNI allows direct access to 3D data, which is very helpful in developing natural interaction-based applications such as motion tracking and activity analysis [8].

3.2 Strategies to Avoid Opponents Using Distance Data

Opponent avoidance strategies are an important element in the navigation system of wheeled soccer robots, such as those used in KRSBI-B. By utilizing stereo vision cameras, robots are able to estimate the actual distance between themselves and opposing robots that are blocking the path to targets such as the ball or goal. When the system detects that an opponent is in the main path and the distance is below the safety threshold (e.g., < 1 m), the robot automatically takes alternative maneuvers to avoid collision. These maneuvers can be lateral shifts (left or right), turns, or circular routes adjusted to the relative position of the opponent and the main destination direction [16]. This real-distance-based approach is applied in wheeled robots that utilize YOLO and stereo depth maps to determine safe paths in real time and has been successfully executed on wheeled soccer robots on the field [6]. The equation can be formulated as follows :

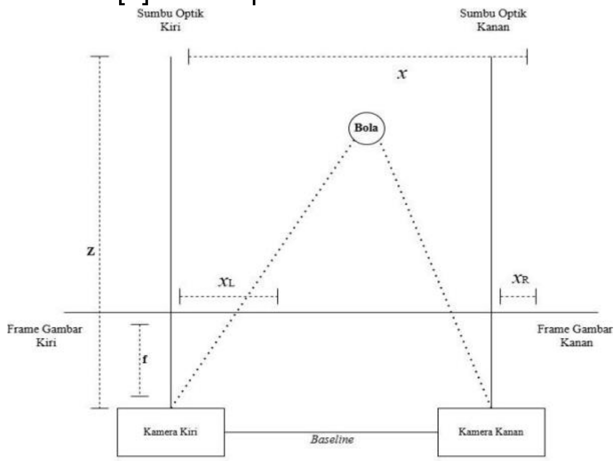


Fig. 2. Stereo Vision Camera Disparity System

$$\frac{Z}{f} = \frac{X}{Xl} \tag{3.1}$$

$$\frac{Z}{f} = \frac{X - b}{X_R} \tag{3.2}$$

$$X_l = \frac{X}{Z}\ x\ f \tag{3.3}$$

$$X_R = \frac{X - b}{Z}\ x\ f \tag{3.4}$$

$$Disparity = X_L - X_R \tag{3.5}$$

- Z: the estimated depth or real distance of the object from the camera.
- f: the focal length of the stereo camera.
- b: the baseline distance between the left and right cameras.
- X: the horizontal position of the object in three-dimensional space.
- X_L: the horizontal position of the object detected in the left camera image.
- X_R: the horizontal position of the object detected in the right camera image.
- **Disparity ($X_L - X_R$)** : the difference between the object positions in the left and right images, which is used to estimate the depth of the object.

The disparity system in stereo vision cameras is used to estimate the depth or distance of objects by exploiting the difference in viewpoints between two cameras separated by a fixed distance, known as the baseline. In this formulation, Xrepresents the horizontal position of the object in three-dimensional space, while $X_L$and $X_R$denote the horizontal positions of the object detected in the left and right camera images, respectively. The variable Z represents the depth or distance of the object from the camera, $f$is the focal length, and $b$is the baseline distance between the two cameras.

Equations (3.1) and (3.2) describe the relationship between the object depth and its projection on the left and right camera images. Equations (3.3) and (3.4) compute the image coordinates of the object based on its spatial position and depth. Finally, Equation (3.5) defines the disparity as the difference between the object positions in the left and right images. A larger disparity value indicates that the object is closer to the camera, while a smaller disparity corresponds to a farther object. The estimated depth information obtained from the disparity is then used by the navigation system to support opponent avoidance decisions and safe path planning, including the detection of balls and opponent robots on the field.

3.3 Harris Corner Detection

A corner can be defined as a point that has a local environment consisting of two different and dominant edge directions called corners. An angle can be described as the intersection of two sides with a sharp change in brightness intensity in an image. Harris Corner Detection is an effective method for detecting angles, as it is able to distinguish between edges and angles with high accuracy. This method produces consistent values even when the image undergoes rotation, scaling, lighting variations, or noise interference.

The angle detection process by Harris Corner Detection is performed by calculating the variation in intensity values in the image using a binary window that is shifted along the x and y axes. To determine the variation in intensity values, a specific equation can be used. To determine the variation in intensity values, equation 2.1 can be used to calculate the change in intensity in the image.

$$
\begin{aligned}
E(\mathrm{u},\mathrm{v}) &= \sum_{\mathrm{x,y}} \mathrm{w}(\mathrm{x},\mathrm{y})[\mathrm{I}(\mathrm{x}+\mathrm{u},\mathrm{y}+\mathrm{v}) - \mathrm{I}(\mathrm{x},\mathrm{y})]^2 \\
&= \sum \mathrm{u}^2\,\mathrm{I}_{\mathrm{x}}^2 + 2\mathrm{uv}\mathrm{I}_x\mathrm{I}_y + v^2 I_y^2 \\
&= \sum[\mathrm{u}\ \mathrm{v}]\left(\begin{bmatrix} \mathrm{I}_{\mathrm{x}}^2 & \mathrm{I_xI_y} \\ \mathrm{I_xI_y} & \mathrm{I}_{\mathrm{y}}^2 \end{bmatrix}\right)\begin{bmatrix}\mathrm{u}\\ \mathrm{v}\end{bmatrix} \\
&= [\mathrm{u}\ \mathrm{v}]\left(\sum\begin{bmatrix} \mathrm{I}_{\mathrm{x}}^2 & \mathrm{I_xI_y} \\ \mathrm{I_xI_y} & \mathrm{I}_{\mathrm{y}}^2 \end{bmatrix}\right)\begin{bmatrix}\mathrm{u}\\ \mathrm{v}\end{bmatrix}
\end{aligned}
\tag{2.1}
$$

- E = is the difference between the original window and the shifted window.
- u = is the shift in the x-axis direction.
- v = is the shift in the y-axis direction.
- w(x,y) = window function to determine the limits on the x-axis and y-axis
- I(x+u,y+v) = the intensity of the shift in the window.
- I(x,y) = the original intensity.

3.4 Object Detection using YOLO as a Basic for Strategy Determination

You Only Look Once (YOLO) is an algorithm derived from a Convolutional Neural Network (CNN) designed to help people find objects in real time. This algorithm can produce up to 45 frames per second. The YOLO detection system uses a classifier or localizer to find objects in various locations and scales. The area of the image with the object will receive the highest score, indicating that the object has been successfully found [5].

YOLOv5 is one of the latest versions of the You Only Look Once (YOLO) algorithm, which is better than previous versions. This version offers a better anchor system, detection capabilities on various scales, and higher efficiency in the backbone network structure [6].

YOLOv5 was chosen because it has a lightweight architecture and uses a single-stage detection method, which allows high processing speed while maintaining stable accuracy. This is very important for wheeled soccer robot competitions that require fast and dynamic responses.

3.5 Robot Operating System (ROS)

The Robot Operating System (ROS) is a software framework that runs on the Linux operating system and is specifically developed to support the development of mobile robots [7]. ROS provides facilities for developers to design, test, and manage complex robotics systems. The ROS architecture consists of a collection of nodes, where each node is an independent program that performs a specific function. Communication between nodes is coordinated by the ROS Master, which functions as a message exchange center. With this structure, ROS enables the creation of integrated robotics systems from various components. In addition, ROS is equipped with a variety of visualization tools to monitor and analyze robot performance [17]. This platform can be used on various types of robots, both mobile robots and manipulator robots, and is compatible with a number of operating systems such as Linux, Windows, and macOS.

Fig. 3. Robot Operating System Logo

ROS is used as middleware because it supports real-time communication between system modules. This allows efficient and modular synchronization of vision data, distance estimation, and robot motion control.

3.6 Hardware Implementation

The mechanical design of the wheeled soccer robot was created with functionality, stability, and Pusprenas 2024 regulations in mind. The design was created using Autodesk Fusion 360 with lightweight materials such as aluminum to keep the weight below 40 kg. The robot uses omni or mecanum wheels for agile maneuvering. The ball-following system uses rubber rollers, while the kicker uses a solenoid actuator for fast kicks. The components are designed to be modular for easy assembly and maintenance, as well as to support optimal performance during matches.

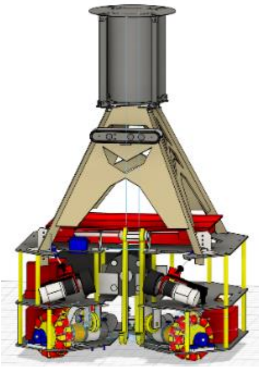


Fig 4. Front View of Robot Design

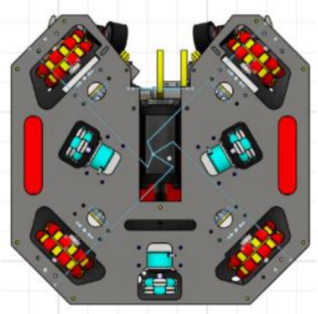


Fig 5. Bottom Views of Robot Design

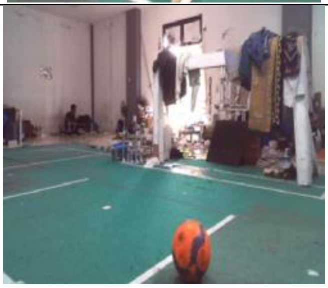
Fig 6. Robot Implementation

4. Experimental Setup

4.1 YOLO Model Training Dataset

The dataset used to train the YOLOv5 model focused specifically on one class of objects, namely "robots," which were identified based on their distinctive visual characteristics: black color and upward-tapered shape. To build this dataset, images of robots were collected and labeled manually. Next, data augmentation processes were carried out, such as rotation, scaling, and lighting and contrast adjustments to enrich data variation and make the model more resilient to various conditions in the field. The reliability of the model was then tested in difficult scenarios, such as when the robot was obstructed, moving, or in unstable lighting, and validated with separate data to prevent overfitting. The ultimate goal of preparing this structured dataset was to create a fast and accurate robot detection system so that wheeled soccer robots could respond optimally during matches.

Table 1. Ball and Robot Dataset

Class	Dataset	Total
Robot		2000
Bola		2000

4.2 Dataset Pre-processing Stage

The YOLO object detection model was evaluated after undergoing a training process using a dataset containing 4,000 images with one main class, namely robots. Each robot object in the image was labeled manually so that the model could learn to recognize its position, which is an important aspect in supporting competition strategies on the field. In this study, we divide the training data (70%), with a total of 2,800 images, which were used in the training process so that the model could learn the patterns, features, and visual characteristics of the robot objects. We also provide validation data (20%), with a total of 800 images, which were used to monitor the model's performance during training. This data plays an important role in hyperparameter adjustment and prevents overfitting, which is a condition where the model only memorizes the training data without being able to work well on new data. Test data (Test – 10%) with a total of 400 images that have never been used before are provided as final evaluation material. This subset serves to measure the model's performance in dealing with real data and assess its generalization capabilities. It is expected to not only have high accuracy on training data, but also be reliable and adaptable to new images in the competition environment.

4.3 Strategy Simulation at the Gazebo

Gazebo is open-source simulation software designed to simulate hardware or dynamic systems, especially in the context of robotics. Developed by the Open Source Robotics Foundation, Gazebo allows users to design, test, and simulate robots in various conditions and environments [10]. With its highly accurate simulation capabilities, Gazebo allows users to virtually model robots and other hardware, as well as train artificial intelligence (AI) systems to interact with the environment. Virtual environment simulation using Gazebo is carried out to test the results of path planning using real distance estimates. This environment is designed to resemble the actual field conditions at KRSBI-B, with robot and goalpost models adjusted in size and scale. The purpose of the simulation is to observe the robot's behavior when following the planned path and to ensure that the system works optimally before being physically implemented.

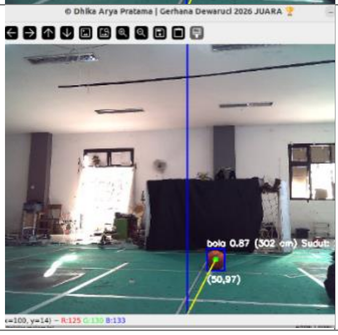
Fig 7. Gazebo Simulation

5. Result and Analysis

In this study, we conduct accuracy testing of detection and distance estimation using the Orbbec Astra stereo camera sensor to assess the camera's ability to accurately detect opposing robots. The robot was placed at a certain distance starting from 3 meters, then measured with a meter. This test involved parameters such as color, ball shape, detection confidence level, as well as object position and distance, which were compared with physical measurements. The Orbbec Astra stereo camera detected objects in the image frame with red lines marking the boundaries of the object and background. The distance estimated by the camera was compared to the physical distance to evaluate detection accuracy. Table 2 describes Ball Object Detection Testing.

Table 2. Result of Ball Object Detection Testing

No	Picture	X Ball	Y Ball	Confidence	Actual Distan (cm)	Camera Distance (cm)	Erorr %	Lux
1		11	39	0.96	50	57	14%	**220**

2		-10	127	0.98	100	106	6%	**231**
3		-2	186	0.98	150	155	3.33 %	**260**
4		-13	216	0.98	200	205	2.5%	**285**
5		-13	272	0.99	250	259	3.6%	**305**
6		-12	250	0.97	300	319	6.33 %	**333**

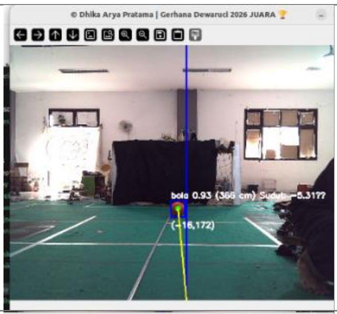
7		-13	261	0.96	350	366	4.57 %	**341**
8		-12	100	0.28	400	419	4.75 %	**359**
9		-2	106	0.48	450	458	1.78 %	**302**
10		**4**	**160**	**0.45**	**500**	**505**	**1%**	**312**

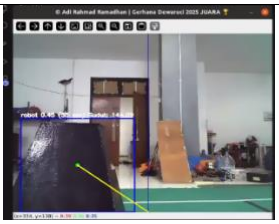
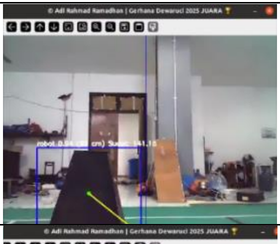
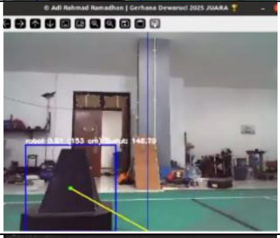
Table 2 presents the results of spherical (ball) object detection and distance estimation using the stereo vision camera. The results indicate that the system exhibits excellent detection performance at close to medium distances, with confidence values approaching 1.0 for distances up to 350 cm. This demonstrates that the YOLOv5-based detection model is capable of identifying ball objects with high precision under adequate lighting conditions and within the effective operating range of the camera.

As the object distance increases beyond 350 cm, a gradual decrease in detection confidence is observed, with confidence values dropping to 0.28 at 400 cm and 0.45 at 500 cm. This reduction in confidence indicates that object visibility decreases at longer distances, which may be caused by reduced image resolution, depth sensing limitations, and environmental lighting variations.

In terms of distance estimation, the stereo camera generally produces distance values that are slightly higher than the actual measured distances. For example, at an actual distance of 50 cm, the estimated camera distance is 57 cm, while at 500 cm, the estimated distance is 505 cm. Although small deviations exist, the difference between the actual distance and the estimated distance remains relatively consistent across the tested range.

Based on the corrected percentage error calculation, the distance estimation error for ball objects ranges between approximately 1% and 6% for most measurements, indicating that the proposed stereo vision system provides stable and reliable distance estimation performance. These results confirm that the system is suitable for real-time ball tracking and distance-based strategy implementation in wheeled soccer robot applications.

Table 3. Robot Object Detection Test Result

No	Picture	X Robot	Y Robot	Confidence	Actual Distance (cm)	Camera Distance (cm)	Error %	Lux
1		35	125	0.45	50	55	10%	22
2		20	121	0.94	100	99	1%	25
3		27	128	0.91	150	153	2%	26
4		29	146	0.94	200	205	2.5%	24
5		26	160	0.80	253	267	5.53%	29

6		19	165	0.88	300	315	5%	30

Table 3 presents the experimental results of opponent robot detection and distance estimation using the stereo vision camera. The results show that the estimated distances obtained from the camera are generally close to the actual measured distances, with only small deviations observed across all test samples. For instance, at an actual distance of 100 cm, the camera estimates a distance of 99 cm, indicating a very small measurement error.

In general, the stereo camera tends to slightly overestimate the actual distance, particularly at close and medium ranges. However, the magnitude of this deviation remains relatively small and consistent, making the estimated distance values acceptable for robot object detection and navigation purposes. Based on the corrected percentage error calculation, the distance estimation error for robot objects ranges from approximately 1% to 10%, with the highest error occurring at very close distances (50 cm). At medium distances between 150 cm and 300 cm, the error remains below 6%, indicating stable distance estimation performance within the effective operational range of the system.

The detection confidence values for robot objects are generally high, especially at medium distances between 150 cm and 300 cm, where confidence values exceed 0.8. This demonstrates that the YOLOv5-based detection model is capable of reliably identifying opponent robots when their visual features are clearly captured by the camera.

Data analysis in Table 2 shows variations in the accuracy of distance detection between the actual distance and the camera distance for spherical objects. The confidence values generally range from 0.96 to 0.99 for distances up to 350 cm, indicating that the ball object detection system tends to be accurate within this range. However, a noticeable decrease in confidence occurs in the eighth and tenth tests, where the confidence values drop to 0.28 and 0.45, indicating increased uncertainty in ball object detection at longer distances.

The error values vary across the tests, with the highest error percentage reaching approximately 14% in the first test, which indicates a relatively larger discrepancy between the actual distance and the camera distance at a very close range. This higher error is caused by the ball object not being perfectly centered in the image frame, which affects the accuracy of disparity-based distance estimation. Meanwhile, the lowest error is observed at longer distances, with an error of approximately 1%–2%, indicating better distance estimation consistency under stable detection conditions. Overall, the distance estimation error for ball objects ranges from approximately 1% to 6% for most measurements, reflecting moderate variation in accuracy under different testing conditions.

The actual distance tested varies between 50 cm and 500 cm, and the estimated camera distance generally increases as the actual distance increases, with only small deviations between the two. For example, at an actual distance of 50 cm, the camera estimates a distance of 57 cm, while at 500 cm the estimated distance is 505 cm, showing relatively consistent measurement behavior. The lux or lighting level used in these tests ranges from 220 lux to 359 lux, which influences the detection performance of the ball object. Lower lighting conditions tend to reduce detection confidence and increase estimation error, while higher illumination levels produce more stable results. Overall, although the system demonstrates strong potential for ball object detection, improvements in lighting robustness and detection model optimization are required to further enhance the consistency and accuracy of distance measurement results.

This study has several limitations, including reduced distance estimation accuracy under low-light conditions and at long distances. In addition, the opponent avoidance strategy still relies on static threshold values, which are not fully adaptive to the dynamic speed of opponent robots.

6. Conclusion

This study demonstrates the successful implementation of a real-world distance estimation system for wheeled soccer robots by integrating stereo vision cameras with the YOLO-based object detection framework. Experimental evaluations involving 10 ball detection samples and 6 robot scenarios confirm that the proposed system achieves a high average accuracy of 96.7%, with a low estimation error of ±2.3%. The system also attains a confidence level of 99%, indicating strong reliability and consistency in both object detection and metric distance measurement. These results validate the effectiveness of combining stereo disparity computation with deep learning–based detection for supporting perception tasks in dynamic robot soccer environments.

From a practical perspective, the achieved accuracy and reliability directly enhance the robot's ability to execute game strategies that depend on precise spatial awareness, such as positioning, passing, and shooting decisions. Accurate real-distance information allows wheeled soccer robots to respond more effectively to fast-moving objects and opponents, reducing uncertainty in tactical planning. Moreover, the use of stereo vision cameras ensures compliance with competition constraints while maintaining robust depth perception without relying on expensive or restricted depth sensors.

Despite the promising results, several challenges remain that open opportunities for future work. Variations in illumination, shadows, and rapid motion still affect stereo matching quality and detection stability in certain match conditions. Future research will therefore focus on improving robustness under diverse lighting environments, adopting newer and more efficient YOLO model variants, and integrating artificial intelligence–based adaptive strategy modules. By coupling accurate distance estimation with learning-based decision-making, future systems are expected to further improve autonomy, adaptability, and overall performance in wheeled soccer robot competitions.

References

[1] B. Benyamin, I. M. K. Putra, and M. A. Kusumoputro, *Guidelines for the Indonesian Robot Contest (KRI) 2023*, Balai Pengembangan Talenta Indonesia, Pusat Prestasi Nasional, Ministry of Education, Culture, Research, and Technology, Jakarta, Indonesia, 2023.

[2] Y. Hu, G. Liu, Z. Chen, and J. Guo, "Object detection algorithm for wheeled mobile robots based on an improved YOLOv4," *Applied Sciences*, vol. 12, no. 9, Art. no. 4769, 2022, doi: 10.3390/app12094769.

[3] R. Ramadhan, A. Khumaidi, M. K. Mayangsari, M. Syai'in, I. Sutrisno, and A. R. Annisa, "Application of Harris corner detection and YOLOv5 on stereo vision cameras for distance estimation of wheeled soccer robots (KRSBI-B)," *Journal of Electronics and Industrial Automation*, vol. 12, no. 1, pp. 111–122, 2025, doi: 10.33795/elkolind.v12i1.7254.

[4] Biswas, B. Dey, B. Poudyel, N. Sarkar, and T. Olariu, "Automatic fall detection using Orbbec Astra 3D Pro depth images," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 1707–1715, 2022, doi: 10.3233/JIFS-219272.

[5] H. A. Saputri, M. Avrillio, L. Christofer, V. Simanjaya, and I. N. Alam, “Implementation of the YOLOv7 algorithm for traffic flow estimation,” *Procedia Computer Science*, Elsevier, pp. 117–126, 2024, doi: 10.1016/j.procs.2024.10.235.

[6] Wahyudi, A. Khumaidi, M. B. Rahmat, D. P. Riananda, M. Syai’in, and J. Endrasmono, “Implementation of the Robot Operating System to improve ball detection accuracy using YOLOv5 on wheeled soccer robots,” *Journal of Electronics and Industrial Automation*, vol. 11, no. 2, pp. 648–661, 2024.

[7] F. Li and P. Li, “Computer-aided teaching software for three-dimensional motion models based on Kinect depth data,” *Computer-Aided Design and Applications*, vol. 18, no. S2, pp. 123–134, 2020.

[8] M. Szemenyei and V. Estivill-Castro, “Fully neural object detection solutions for robot soccer,” *Neural Computing and Applications*, vol. 34, pp. 21419–21432, 2022, doi: 10.1007/s00521-021-05972-1.

[9] W. Darmawan, M. B. Rahmat, A. Khumaidi, R. Y. Adhitya, and D. P. Riananda, “Decision-making strategy design for wheeled soccer robots using decision tree methods,” *Journal of Electronics and Industrial Automation*, vol. 10, no. 2, pp. 175–182, 2023.

[10] H. Kusuma, D. Ahmad, and R. D. Department, “Assisted depth imaging for scoring goals in wheeled soccer robots,” *IEEE Access*, vol. 11, pp. 112345–112356, 2023.

[11] Winarti, Susanto, R. Analia, and E. R. Jamzuri, “Improving stereo distance measurement accuracy on humanoid soccer robots,” in *Proc. IEEE Int. Conf. Applied Engineering (ICAE)*, 2022, doi: 10.4108/eai.5-10-2022.2327753.

[12] Khumaidi, “Soccer robot position mapping using gyrodometry and trigonometry for kick angle prediction,” *Poly-Technology Journal*, vol. 19, no. 3, pp. 271–277, 2021, doi: 10.32722/pt.v19i3.2864.

[13] Abdelsalam, M. Mansour, J. Porras, and A. Happonen, “Depth accuracy analysis of the ZED 2i stereo camera in indoor environments,” *Robotics and Autonomous Systems*, vol. 175, Art. no. 104753, 2024, doi: 10.1016/j.robot.2024.104753.

[14] McManus, W. Churchill, W. Maddern, A. D. Stewart, and P. Newman, “Shady dealings: Robust, real-time stereo vision in challenging illumination,” in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2013, pp. 1162–1167.

[15] S. H. Abood, H. M. H. Al-Khafaji, and M. M. H. Al-Khafaji, “Enhancing collision avoidance in mobile robots using YOLOv5,” *Journal of Robotics and Control*, vol. 6, no. 2, pp. 769–778, 2025, doi: 10.18196/jrc.v6i2.25856.

[16] S. Nalpantidis and A. Gasteratos, “Stereo vision depth estimation methods for robotic applications,” in *Depth Map and 3D Imaging Applications*, IGI Global, pp. 397–417, 2011.

[17] J. Z. Kolter, Y. Kim, and A. Y. Ng, “Stereo vision and terrain modeling for legged robots,” in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2009.

[18] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, 2004.

[19] Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International Journal of Computer Vision*, vol. 47, no. 1–3, pp. 7–42, 2002.

[20] K. Konolige, M. Agrawal, and J. Sola, “Large-scale visual odometry for rough terrain,” in *Proc. International Symposium on Robotics Research*, Springer, 2007, pp. 201–212.

[21] 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, 2018.