

Research on Image-Guided Positioning Technology for Automated Agricultural Machinery Based on Convolutional Neural Networks (CNN) and Visual Servo Control

Jian Yan

School of Automotive Engineering, Chongqing Three Gorges Vocational College, Chongqing 404155, China

Corresponding author: Jian Yan (e-mail: yanjian6666@163.com)

ABSTRACT Automated agricultural machinery often suffers from low positioning accuracy and unstable target recognition in complex farmland environments. These problems affect path planning, operation precision, and the reliability of autonomous control, especially when visual perception must operate alongside wireless communication and electromagnetic sensing modules in harsh outdoor conditions. This paper addresses these issues by constructing an image-guided positioning model based on convolutional neural networks (CNNs) and visual servo control. The model acquires crop and environmental images through a multi-source image acquisition module, extracts features and recognizes targets by an improved lightweight CNN with an attention mechanism, and inputs the recognition results into a visual servo control system. Real-time position correction is performed together with depth-sensor feedback to achieve accurate positioning and path tracking. The method includes establishing a multi-scale image sample set, designing an attention-enhanced CNN, embedding proportional-integral-derivative adjustment into the visual servo loop, and correcting system deviations through an error-compensation model. Experimental results show that the positioning error remains between 2.4 cm and 2.8 cm in complex field environments, with an average response speed of 83.8 ms. The method therefore improves the automation, stability, and engineering applicability of image-guided agricultural machinery.

INDEX TERMS convolutional neural network, visual servo control, image-guided positioning, electromagnetic sensing, autonomous machinery

I. INTRODUCTION

AUTOMATED agricultural machinery demonstrates significant application value in precision agriculture. The technical necessity of integrating high-performance industrial fiber-reinforced structures lies in providing the mechanical damping and shielding required for sensitive electronic components; specifically, these textile-based composites stabilize the visual sensor housing against high-frequency vibrations during soil interaction, thereby ensuring the data integrity of the CNN input. By embedding advanced fiber-reinforced composite structures into intelligent platforms, these systems achieve the structural durability and data-gathering precision necessary for autonomous agricultural operations. However, in complex field environments, its operation is still constrained by problems such as insufficient positioning accuracy and unstable visual recognition, affecting the planning of operation paths and operational efficiency. Achieving high-precision, adaptive image-guided positioning has become the key to improving the intelligence

level of agricultural machinery [1,2]. This study focuses on visual perception and combines convolutional neural networks (CNN) and visual servo control technology to explore stable recognition and precise control methods in complex lighting, unstructured terrain, and diverse crop environments, so as to promote the development of agricultural automation technology.

Domestic research on visual navigation of agricultural machinery mainly focuses on the optimization of image processing and path planning algorithms. However, due to the limitations of algorithm generalization ability and environmental adaptability, the actual field operation effect is still insufficient. Foreign research introduced deep learning and visual servoing methods earlier, which improved the perception accuracy and control stability. However, the system mostly relies on high-cost sensing equipment, which is difficult to promote and apply in large-scale agriculture. Therefore, developing an image-guided positioning technology for agricultural machinery based on lightweight CNN

and high robustness servo control has significant theoretical and practical significance [3].

This study addresses the core problem of positioning and control of agricultural machinery in unstructured environments, proposing a joint control method that integrates visual servo feedback and deep feature recognition. The research includes multi-source image data acquisition and preprocessing, construction of a CNN model with an attention mechanism, optimization of the servo system based on proportional-integral-derivative (PID) control, and design of an error compensation model [4,5]. Technical innovations center on optimizing a lightweight feature extraction network and dynamic visual servo loop, incorporating high-strength industrial fiber-reinforced sensors to stabilize the agricultural machinery's structural response during high-precision image-guided positioning. By fusing textile-based flexible sensing with the adaptive adjustment system, this low-cost framework achieves superior data acquisition and positioning accuracy within complex field environments.

II. THEORETICAL FOUNDATIONS AND SYSTEM DESIGN

Visual servo control is a control method that uses visual sensors to acquire environmental image information and adjust the motion of actuators accordingly. Its theoretical basis is the mathematical mapping between image features and pose changes. This control is generally divided into position-based visual servo (PBVS) and image-based visual servo (IBVS). This paper adopts the IBVS structure, directly using image error signals to drive the controller output, and dynamically corrects the pose through a proportional-integral-derivative (PID) algorithm to achieve real-time, high-precision mechanical motion control.

The systemic layout of the image-guided positioning system consists of four primary modules, which are an image acquisition module, a feature recognition and positioning module, a visual servo control module, and an actuator module. This system obtains first the images of the target crop and terrain by using multi-source visual sensors. The target position coordinates are then retrieved after preprocessing the image and extraction of features with a CNN model. The attitude correction is computed by the visual servo controller and the underlying actuator is used to carry out the displacement operation [6,7]. This mechanism also provides a closed-loop feedback architecture of image perception to mechanical control, so that mechanical farm equipment is automatically detected, positioned and tracked.

The data acquisition and multi-source information fusion scheme will be to enhance the ability of the system to adapt to the complicated environment of the farmlands. Multi-dimensional data are obtained by the combination of visible light cameras, depth cameras, and inertial measurement units (IMUs) and are temporally and spatially aligned with the help of a calibration model. This is followed by a process of fusion, using an Extended Kalman Filter (EKF) framework to integrate asynchronous data streams. By defining a state-space model that couples visual feature

vectors with IMU-derived acceleration and depth-based spatial constraints, the system resolves pose estimation across varying time-stamps to create high-precision spatial point cloud representation [8,9]. The result of the fusion, once fed into a CNN model, can be useful in reducing the difference in illumination and noise interference as well as improving the perception abilities of the system in moving environments and supplying the data with stable values to support accurate positioning and pathfinding.

III. DESIGN AND IMPLEMENTATION OF IMAGE-GUIDED LOCALIZATION ALGORITHM

A. IMAGE FEATURE MODELING AND SAMPLE SET CONSTRUCTION FOR AGRICULTURAL SCENES

Multi-source platform of image acquisition completes the feature modeling of agricultural scene images. The system has a high-resolution RGB camera and a depth camera to obtain surface texture and spatial geometric data at the same time. In the process of acquisition, the time sequence control module achieves image frame synchronization and image brightness and color consistency processing, which is performed by the illumination equalization algorithm on the data in each time period. In the case of the characteristics of the crop leaf texture, inter-row soil structure, and weed distribution, a morphological description matrix is developed with the weighted gradient histogram and spatial scale normalization. The CNN input feature benchmark in the network training is this matrix. The sample set construction phase uses the semantic mask-based regional labeling strategy and produces pixel-level labels created using high-precision labeling tools in differentiating target crops and the areas containing interest in the interference background. Information supplementation embraces random affine transformation and spectral perturbation to preserve the continuity of texture and enhance the distribution variety of data [10]. Each type of crop has an equal number of samples within the network to make sure that the network has consistent detection and recognition in various landform environments.

B. LIGHTWEIGHT CNN MODEL DESIGN INCORPORATING ATTENTION MECHANISM

The model structure integrates separable convolution depth-wise and channel compression methods to minimize computational complexity. Quantitatively, the proposed lightweight architecture achieves a parameter count of 4.2M and a computational load of 0.85 GFLOPS, representing a 65% reduction in complexity compared to the standard ResNet-18 baseline while preserving the quality of feature representation. The farmland picture data are taken at three channels in the input layer, and the local texture value is extracted using 5×5 convolutional kernels, and this is followed by the output being subjected to the normalization of the batch to stabilize the gradient distribution. Shallow feature maps have a concatenated feature map with a deep semantic feature having a bottleneck structure that lowers

the number of parameters. The network has a channel attention module, and the global average pooling results are taken as weight inputs. Channel importance is rearranged by one-dimensional convolution, boosting crop region feature response whilst reducing environmental background. Spatial attention module enables use of convolution to combine high-response areas in the feature map in a global spatial weight distribution, which enhances the sensitivity of target edge localization. The output layer in the model employing the Softmax classification function converts the feature mapping to probability matrices of crop, soil and obstacles. It further integrates the loss of cross-entropy to optimize parameter step to allow the network to possess efficient recognition and real-time computing performance in complex farmland environments.

C. OPTIMIZATION OF VISUAL SERVO CONTROL LOOP AND PID ALGORITHM

The image feature error feedback mechanism forms the basis of the visual servo control loop. The system generates a semantic mask for the crop area; it then calculates the geometric center of the segmented pixel cluster to define the target reference point within the image coordinate system and derives the deviation vector about the ideal path. The control law makes use of a pixel displacement mapping model that is founded on the Jacobian matrix that transforms the two-dimensional image error into a motion control quantity of the robotic arm end effector in three-dimensional space. The system comprises of an image sensor, error estimation module, PID controller and execution unit in a loop. The control signal is revised with every servo cycle so that the response to the positioning of the agricultural machinery is maintained at a steady level. A dynamic mechanism of weight adjustment is also proposed by the PID control algorithm, which is adjusted adaptively according to the actual error rate and its varying tendency to achieve the smoothness and quick convergence of the control process in the non-linear terrain. The integral channel employs a moving average filter to eliminate cumulative drift and the derivative channel adds a discrete Laplace operator to minimize noise interference and overshoot response.

Table 1, Δe represents the change in image error, $\sum e$ is the cumulative error term, and $\Delta^2 e$ is the second-order error change term. The smoothness and sensitivity of the α and β control loops are adjusted together to achieve posture stability and continuous path movement of agricultural machinery in complex terrain.

D. POSITIONING ERROR COMPENSATION AND MODEL PARAMETER ADJUSTMENT

The key payoff in positioning error compensation process is based on the continuous pose data correction and network parameter fine-tuning by means of multi-source sensor signal fusion to deliver dynamic correction of agricultural machinery position. The system models an error distribution with historical path residues, and does weighting estimation

of pose offset at every moment. The outputs of the vision module, pixel positions are spatially transformed and combined with angular velocity sensor output of the Inertial Measurement Unit (IMU) as error correction input. The compensation module breaks the error components down into two independent directions, translation and rotation and computes the correction magnitude of each and feeds it back to the execution control layer synchronously. The strategy of model parameter adjustment uses an empirical update strategy, dynamically assigning adjustment weights to network and control parameters depending on the stability of the environment, the intensity of illumination, and surface parallax. A sliding window is employed by the system to assess the short-term and long-term trends of errors, making sure that the performance of positioning does not exceed a stable convergence margin.

Table 2 are monitored in real time during system operation, and the update frequency of different modules is dynamically adjusted according to stability. This strategy achieves collaborative compensation between the vision and control layers, enabling agricultural machinery to maintain high-precision trajectory execution capabilities even under complex terrain and lighting interference.

IV. RESULTS AND DISCUSSION

A. CONSTRUCTION OF EXPERIMENTAL PLATFORM AND TESTING ENVIRONMENT

The prototype intelligent agricultural machinery, binocular vision camera system, depth sensing module and embedded computing unit are the components of the experimental platform. The machinery platform will have a four-wheel drive chassis, which is installed on a high-precision steering serving mechanism and a position feedback encoder to determine the viability of image-guided positioning. The vision system consists of a binocular camera with a resolution of 1280×720 at a baseline distance of 60 mm along with infrared depth camera to have a synchronous data acquisition. The onboard computer is based on an NVIDIA Jetson Orin NX development board, with an operating system of Ubuntu 20.04 and the PyTorch deep learning system, utilizing CUDA to speed up the calculation of convolutional neural network models. The experimental conditions are configured in an artificial farmland simulation area, where there are various forms of crops, light conditions, and unevenness of the ground so that the experimental conditions may cover a wide range of conditions. The positioning error is measured with a Leica TS16 high-accuracy total station (precision of $1 \text{ mm} + 1.5 \text{ ppm}$) to calibrate the ground coordinates. The machine's internal coordinate system is synchronized with the global reference frame through a three-point resection method, establishing a common spatial origin to ensure the validity of the 2.4 cm error margin. Control signals and image streams are sent in real time by means of the ROS communication system. To determine the stability and consistency of the response of the algorithm with varying operating speeds, the system data sampling

TABLE 1. PID parameter adjustment strategy for visual servo control loop

Parameter	Initial value	Correction coefficient	Update conditions	Function Description
P	0.85	$0.03\Delta e$		Adjusting the proportional response speed
I	0.12	$0.05 \sum e$		Correcting steady-state error
D	0.06	$0.02\Delta^2 e$		Suppressing error oscillations
α	0.9	Dynamic adjustment		Filtering and smoothing weights
β	0.5	Fixed coefficient		Control output suppression coefficient

TABLE 2. Explanation of Positioning Error Compensation and Parameter Adjustment Strategies

Adjustment phase	Monitoring indicators	Adjustment basis	Update frequency	Function Description
Image feature matching	Pixel displacement difference	Texture clarity and edge alignment	Every 20 frames	Correction of visual error accumulation
Deep fusion correction	Spatial residual mean	Deep point cloud stability	Every 10 frames	Improve 3D positioning consistency
Inertial compensation	attitude drift angle	IMU angular velocity drift amplitude	Every 5 frames	Stabilizing mechanical attitude changes
Learning rate balance	Network gradient rate of change	Ambient lighting and range of motion	Every 50 frames	Dynamically control the convergence speed of CNN parameters
Control gain fine-tuning	Execution delay time	Control signal response stability	Each cycle	Maintaining path tracking continuity

frequency will be 30 Hz. Under the NVIDIA Jetson Orin NX environment, the optimized CNN model achieves an average inference latency of 12.4 ms per frame. This ensures that the perception-to-control pipeline is completed well within the 33.3 ms sampling interval, preventing servo loop lag and maintaining real-time synchronization.

B. EXPERIMENTAL RESULTS OF POSITIONING ACCURACY AND RESPONSE SPEED

The accuracy of positioning and speed of response experiments were carried out on the proposed system and compared against both the traditional Feature Matching (FM) method and a modern deep-learning-based baseline, MobileNetV3-SSD. The experiments were conducted under the same field and mechanical platform. The positioning error was determined through comparing the total station coordinates and the speed of the system response was determined by the time difference between transmissibility of the control signal and the execution response. A test consisted of three phases, namely path detection, target acquisition, and trajectory correction. The mean error and response time of the system in real-time motion were taken on board. The test results are shown in Figure 1.

While the MobileNetV3-SSD baseline achieved an average error of 4.1 cm, the proposed CNN+VSC method consistently maintained a positioning error between 2.4 cm and 2.8 cm, reflecting the superior stability of the integrated attention mechanism.

C. DISCUSSION OF RESULTS

The Convolutional Neural Networks (CNNs) image-guided localization algorithm, with the help of Visual Servo Control, is far much better than the control algorithm concerning the recognition accuracy and speed of response. This is

the performance enhancement that is mainly due to the streamlined network model structure and the synchronized impact of the visual control closed loop. The CNN used is lightweight and optimizes the feature representation of the crop region by using an attention mechanism during the extraction of deep features, so the noise interference of illumination and occlusion is minimized so that the target can be recognized consistently even in complicated settings. The adaptative parameter change and error feedback that is dynamic characteristics of the visual servo loop allow the provision of a lag in the response of the actuator, and thus ensure a quick synchronization of continuous localization and pose correction.

V. CONCLUSION

This paper develops a visual-localization algorithm based on convolutional neural networks and visual servo control, integrating high-precision industrial fiber-optic sensors to enhance the feedback loop for dynamic mechanical responses. By fusing textile-based flexible sensing with intelligent control frameworks, the system achieves a comprehensive level of visual recognition and adaptive positioning accuracy in complex operational environments. This allows the agricultural machines to locate themselves in complicated surroundings with accuracy and independently locate themselves. The enhanced lightweight CNN increases the feature representation saliency based on an attention mechanism, in which the recognition of images can be strongly favored by the servo control process. The visual servo system, along with the feedback and error compensation strategies, allow the execution unit to provide the adaptive control and correction of the system that increases the continuity of the work path and coordination of control. The general model illustrates great robustness and excellent computa-

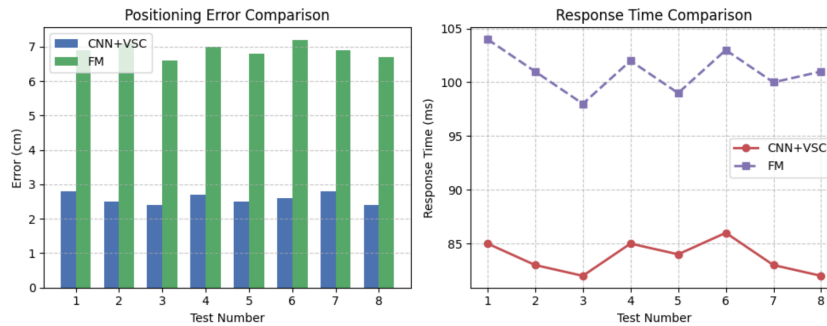


FIGURE 1. Experimental Results

tional performances in the real-life implementation, which can confirm the general feasibility of the coupled visual perception and control architecture within the automation of agriculture. Further research must enhance multispectral image information fusion and low-power embedded deployment, particularly by integrating high-precision industrial fiber-optic sensing to stabilize data acquisition in non-stationary environments. By fusing textile-based flexible electronics with advanced perception models, the system achieves the structural resilience and multi-source data consistency necessary to increase the generalizability of agricultural machinery applications.

AUTHOR CONTRIBUTIONS

Jian Yan designed, collected and analyzed the data, and drafted the manuscript. Jian Yan conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Jian Yan participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research was supported by, Design and Research on a Weeding Machine for Orchard Ridge-shaped Plots in Hilly and Mountainous Areas (KJQN202503522); Design and Research of a Lightweight and Simplified Ancient Red Tangerine Harvesting Device (KJQN202303501).

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] K. Liang, "Design of Automated Agricultural Machinery Chassis Based on Simulink," *Southern Agricultural Machinery*, vol. 54, no. 14, pp. 17-20, 2023, DOI: 10.3969/j.issn.1672-3872.2023.14.005.
- [2] J. Wang, "A brief discussion on the application of intelligent agricultural machinery and automated agricultural production systems," *Guangdong Sericulture*, vol. 58, no. 8, pp. 13-15, 2024, DOI: 10.3969/j.issn.2095-1205.2024.08.04.
- [3] Y. Deng, "Application of automated agricultural machinery in modern agriculture," *Agricultural Engineering Technology*, vol. 44, no. 17, pp. 71-72, 2024, DOI: 10.16815/j.cnki.11-5436/s.2024.17.029.
- [4] D. Lu, X. Xue, Y. Wang, Y. He, and J. Lu, "Design and Implementation of Digital Twin System for Robotic Arm in Automated Agriculture," *Ningxia Agricultural and Forestry Science and Technology*, vol. 65, no. 9, pp. 5-10, 2024, DOI: 10.3969/j.issn.1002-204x.2024.09.003.
- [5] J. Fang, E. Guan, H. Liu, J. Guo, X. Tan, Y. Liu, Y. Zeng, F. Zeng, and Z. Shao, "Effects of different automated agricultural machinery planting methods on the yield and economic benefits of ratooning rice," *China Agricultural Technology Extension*, vol. 41, no. 2, pp. 50-53, 2025, DOI: 10.3969/j.issn.1002-381X.2025.02.016.
- [6] A. Hoque and M. Padhiary, "Automation and AI in precision agriculture: innovations for enhanced crop management and sustainability," *Asian J. Res. Comput. Sci.*, vol. 17, no. 10, pp. 95-109, 2024, DOI: 10.9734/ajrcos/2024/v17i10512.
- [7] S. Sharma, K. Verma, and P. Hardaha, "Implementation of artificial intelligence in agriculture," *Journal of Computational and Cognitive Engineering*, vol. 2, no. 2, pp. 155-162, 2023, DOI: 10.47852/bonviewjccce2202174.
- [8] V. Godase, A. Mulani, S. Takale, et al., "Comprehensive review on automated field irrigation using soil image analysis and IoT," *Journal of Advance Electrical Engineering and Devices*, vol. 3, no. 1, pp. 46-55, 2025.
- [9] J. Lowenberg-DeBoer, Y. Huang I, V. Grigoriadis, et al., "Economics of robots and automation in field crop production," *Precision Agriculture*, vol. 21, no. 2, pp. 278-299, 2020, DOI: 10.1007/s11119-019-09667-5.
- [10] M. Carolan, "Automated agrifood futures: Robotics, labor and the distributive politics of digital agriculture," *The Journal of Peasant Studies*, vol. 47, no. 1, pp. 184-207, 2020, DOI: 10.1080/03066150.2019.1584189.